NATURE AND NURTURE

NATURE AND NURTURE

BEING THE

WILLIAM WITHERING MEMORIAL LECTURES

ON "THE METHODS OF MEDICAL GENETICS" DELIVERED IN THE
FACULTY OF MEDICINE OF THE UNIVERSITY OF
BIRMINGHAM FOR THE YEAR
1933

by

LANCELOT HOGBEN, F.R.S.

*Regius Professor of Natural History in the University
of Aberdeen, formerly Professor of Social
Biology in the University of London*

LONDON
GEORGE ALLEN & UNWIN LTD
MUSEUM STREET

TO

J. B. S. HALDANE

CONTENTS

NATURE AND NURTURE

THE MEDICAL APPLICATIONS OF GENETIC PRINCIPLES

§ 1

Human genetics is a very new department of scientific inquiry. In some quarters comprehensive claims have been put forward on its behalf. One school of opinion holds out the promise that a fuller understanding of the laws of human inheritance can disclose a clue to the rise and fall of civilisations. It is also urged that such knowledge can provide the only basis for a substantial improvement in the common lot of mankind. The grounds for such assertions are open to many criticisms. The basic problem of social evolution is not pre-eminently an investigation of the origin of new types of men and women in a slowly changing physiographical environment. First and foremost it concerns the generation of new modes of behaviour in a rapidly changing man-made environment. In this respect human society has no precise parallel among social organisms. It is a biological phenomenon *sui generis* with unique laws of development. If the laws of its development are ever brought into direct relation to the behaviour of other organisms, it is likely that the study of the central nervous system will have far more to contribute than the study of reproduction. The most formidable problems of civilisation do not arise from limitations in the ability of men and women to command the resources of nature. They arise from imperfect co-ordination of human effort. Experiment alone can decide whether human ingenuity can discover forms of organisation which will guarantee the continued development of our present civilisation. If with its present endowments mankind lacks the capacity to do so, the application of genetic knowledge can only offer a very remote prospect of producing a race which will.

Extravagant assurances of this kind do not alter the fact that human genetics has a genuine claim to be encouraged as a branch of medicine. The study of cancer has very little contribution to

make to a scientific treatment of human history or to the removal
of war, unemployment, and other evils which threaten the stability
of existing civilisation. None the less it is a field of research which
rightly engages public esteem. Whatever else it may accomplish,
genetic science can teach us much about agencies which determine
susceptibility to disease for which curative measures are not yet
available. It may thus make a substantial contribution to the pre-
vention of diseases for which there is no simple remedy. Though it
is not a panacea for human ailments, it is an essential component of
the kind of knowledge upon which scientific prognosis must rest.
Knowledge of this kind is not easy to gain. Special methods are
necessary because of the intrinsic difficulty of dealing with a species
which breeds slowly, has few offspring, and cannot be mated by
the investigator at will. This book is intended to be an intro-
duction to the methods which have been devised already. In this
chapter some contemporary problems will be surveyed. In those
that follow the way in which such problems may be solved will
be examined.

Physicians and their patients often ask such questions as "Should
tuberculous people marry and have children?" or "Are mental
diseases inherited?" Such questions need to be stated in more precise
terms before it is possible to find a precise answer to them. The
difficulty of making them sufficiently precise arises from the fact
that the terms in which the problem of nature and nurture is dis-
cussed are very largely drawn from common speech. An anecdote
in a recent biography of Mendel illustrates this well. It was the
year 1910, and a memorial was being erected to honour Mendel
in the place where he had spent his life teaching and had carried
out his experimental researches. Among the citizens there was
much talk of the distinction which Mendel's discoveries reflected
upon the town. Two visitors were gazing at a portrait of the Abbé
displayed in a shop window. "Do you know who this fellow Mendel
was?" asked one. "Why, yes," was the reply, "he gave Brunn a
bequest (*Vererbung*)."

It has taken scientists a long while to outgrow the confusion
between legal inheritance and biological inheritance. When Darwin
wrote the *Origin of Species* the phenomena of fertilisation had not
been fully elucidated. Biologists still believed that people hand on
their noses to their offspring in much the same way as they hand
on their bank balances. Even now there are biologists of an older

generation who find the prospect of a hundred-per-cent. death duty equally repugnant, whether it is applied to their bank balances or to their noses. Weismann performed a great service to biology by pointing out that the state of death claims all our accumulated anatomical earnings. Our parents do not endow us with characters. They endow us with *genes*. The genes cannot carry their cheque-books into the next life. Weismann has been amply justified by all subsequent work in challenging the belief that the genes have any sympathetic interest in the way we use the organs we owe to them.

Unfortunately, Weismann and his disciples laid too little emphasis on the fact that our genes cannot make bricks without straw. The individual differences which men and women display are partly due to the fact that they receive different genes from their parents and partly due to the fact that the same genes live in different houses. A large body of data dealing with the influence both of nature and of nurture has accumulated since Mendel, Darwin, Weismann, Galton, and their generation first discussed these questions. So we are now able to define what we mean by differences due to genes and differences due to environment with greater precision. That means that we can state in what situations such differences can be distinguished. Before we can ask the right sort of questions about the inheritance of diseases, or, to be more exact, the inheritance of gene differences which affect the liability to contract a disease, we must pause to clarify this distinction with examples from animal genetics.

If chickens are fed on yellow corn or given green food, we can distinguish between some varieties which breed true for yellow shanks and others which breed true for colourless shanks. This is a *genetic difference*. Crosses between such varieties, when all the progeny are fed on yellow corn or given green food, yield numerical ratios of the two types in conformity with Mendel's principle. If chicks of the variety with yellow shanks are fed exclusively on white corn they grow up with colourless shanks. The difference between a fowl of the yellow variety fed on yellow corn and a fowl of the same variety fed on white corn is a *difference due to environment*. If we crossed fowls of the yellow variety with fowls of other varieties, giving some of the progeny yellow corn and others white corn, we could not expect to obtain constant numerical ratios such as Mendel's principle predicts. If two poultry farms, both using yellow corn for food, specialised respectively on birds with black plumage and yellow shanks and birds with barred plumage and

white shanks, we should call both differences genetic differences. If both farms decided to use white corn, we should only be able to recognise the plumage difference as a genetic difference. If both farms varied their procedure quite promiscuously, we should not be able to tell whether the difference between one bird with yellow shanks and another bird with colourless shanks was a genetic difference or a difference due to environment.

Rabbit husbandry provides another illustration of the same issue. Some rabbits deposit yellow fat when fed on green-stuffs. Most rabbits have white fat, whether given greens with their food or not. Yellow fat is a serious carcase defect from a commercial point of view, because purchasers object to it. Rabbits which have white fat when fed on green food possess a liver enzyme which breaks down xanthrophyll, thus preventing it from reaching the fat deposits. Rabbits which deposit yellow fat lack this enzyme. Michael Pease has shown that when rabbits of both kinds are crossed and back-crossed, the absence of the enzyme behaves like an ordinary "recessive character." It is only recognisable as such if the rabbits are given green food containing the yellow pigment. In a group of rabbits of both types we can recognise the gene difference by giving them all green foods. In that case the biological environment is *neutral* and the gene difference is the *isolate* which we are investigating. If none of our rabbits possesses the enzyme which breaks down xanthrophyll, we can make their fat white by feeding them on mash and potatoes, or yellow by feeding them on mash and cabbage. The genetic constitution is then *neutral*, and the biological environment is the *isolate* of the investigation. The practical breeder has therefore two remedies from which to choose. He may put the blame upon the biological environment and cut off the supply of green food. He may put the blame upon heredity and breed for white fat.

In the practice of medicine the same choice may confront us. In some situations the doctor can put the blame for a particular disease on heredity and in others upon environment. An exact biological parallel to *cretinism*, a disease included under the general term *amentia* (idiocy, imbecility and feeble mindedness) illustrates this very clearly. Cretinism turns up occasionally in all communities, and is specially common in certain localities. For this reason doctors sometimes distinguish between a *sporadic* and an *endemic* type of the disease.

In the same way we may distinguish between *genetic* and *ecological* neoteny in Amphibia. Cretinism is due to insufficient quantity of the iodine compound manufactured by the thyroid gland. Insufficiency of the same hormone in Amphibia prevents the aquatic tadpole from transforming into the terrestrial adult. This is sometimes because the thyroid gland is incapable of doing its proper work. Even if it can do so, it cannot make thyroxine without iodine. So if tadpoles are kept in water with no trace of iodine and fed upon a diet free of iodine compounds, they fail to transform into frogs. The European newt normally completes its development and breeds in the adult form. In certain mountainous districts, where *endemic* cretinism is reported among human beings, the newts commonly fail to undergo metamorphosis, or do so after great delay. This is probably because the iodine content of the waters in which they live is low.

A similar explanation does not apply to a local race of the American newt (*Amblystoma tigrinum*) in the neighbourhood of Mexico City. Individuals belonging to this race never grow up; they breed from generation to generation in the aquatic form. They will grow into the terrestrial newt if fed on thyroid gland. They will not do so if given iodine compounds. They possess a thyroid gland which does not release its secretion into the circulation. Failure to undergo metamorphosis in the presence of sufficient iodine sometimes occurs *sporadically* in the Colorado variety among individuals living side by side with others which complete their development. The Mexican variety breeds true for its inability to undergo metamorphosis when kept in aquaria with access to an abundance of iodine compounds.

In the discussion of "mental inheritance" the term *environment* is sometimes equated to training, and even to training at so late a stage as when school education begins. This is very misleading. Even the fact that a condition is congenital provides no presumptive evidence for the view that differences of environment play little part in its occurrence. Lack of thyroid secretion in the maternal circulation may be compared to keeping tadpoles in a tank with iodine-free water and food containing no iodine compounds. At the time of birth a human being has already completed about nine months of its existence as a separate individual. During that time its environment is the womb of its mother, and her physical condition is relevant to the sort of environment in which the most

formative stages of development occur. The fact that a disease is congenital is equally compatible with the belief that genetic differences account for its occurrence, that it is determined by idiosyncrasies of the uterine environment, or that both these agencies play their part in its manifestation. Several things point to the importance of studying the influence of the uterine environment upon the characteristics of individuals. One is the high incidence of certain conditions among first-born children. Another is the high incidence of various malformations among offspring of women approaching the end of the child-bearing period.

A previous example to illustrate the meaning of a genetic difference drew attention to a distinction which is of fundamental importance both for the theory and practice of clinical genetics. In contrasting variations in plumage colour with the colour of the shanks we are not separating a class of phenomena to which the Mendelian principle applies from a class of phenomena to which it does not apply. We are distinguishing between a class of phenomena which are easy to study and a class of phenomena which demand more careful control of the environment. There is no hard-and-fast line between the two. Genetic differences which distinguish plumage colour in fowls are recognisable over a very wide range of environment. This does not mean that they are just as big in every environment which human ingenuity can devise. The difference between the pure black plumage of the Langshan and the mottled plumage of the Light Sussex is a genetic difference. By thyroid feeding, the extent of the black areas in the Light Sussex can be very considerably extended.

In short, no statement about a genetic difference is clear, *unless it includes or implies a specification of the environment in which it manifests itself in a particular manner.* Characteristics of organisms are the result of interaction between a certain genetic equipment contained in the fertilised egg and a certain configuration of extrinsic agencies. The last include the conditions of life in the human uterus and the external environment in which man's social existence is carried on. Differences between individuals may arise from differences in the kind of genes present in the fertilised egg and from differences in the uterine or post-natal environment. Differences of the first kind, that is to say, differences due to a different equipment of genes, may be of two types: (1) differences which are recognisable in almost any environment in which the fertilised egg will develop

and continue to grow; and (2) differences which are only manifest within a fairly restricted range of environment. In the human species examples of the first type are the difference between a haemophiliac (*bleeder*) and an adult whose blood coagulates in the normal way, or between an amaurotic family idiot and an ordinary infant. Medical examples of the second type are less easy to cite because they are less easy to detect. That does not mean that they are less numerous. Probably the best case is the type of mental defect called *Mongolism*. Whatever gene differences are involved in the appearance of this condition appear to require a special pre-natal environment to make them recognisable.

The distinction between the two classes is of the utmost importance from a preventive point of view. When we have to deal with the first, we can readily determine the type of transmission involved, and if we know it, we can estimate the rate at which affected individuals can be eliminated by discouraging parenthood. It is more difficult to determine the method of transmission when a disease belongs to the second class. Unless affected individuals are extremely rare, it is rarely possible to do so, and only so, if we can specify with some precision the kind of environment in which the manifestation of the gene is recognisable. So we cannot give a certain answer to the question: what would be the result of selective interference with parenthood? Usually we could deal with the matter without recourse to selection, if we had the kind of knowledge which tells us how much effect selection would achieve. For instance, we know sufficient to-day about the way in which people get cholera to study the genetic factors involved in susceptibility to the disease among a group of individuals equally exposed to the danger of contracting it. The fact that we have the knowledge to study the problem is the reason why it is of little practical importance to do so. To understand the environmental situation is to be able to control it.

When we understand the *modus operandi* of the gene, we can state the kind of knowledge we need in order to control the conditions in which its presence will be recognised. A variety of the domestic fowl known as the Frizzle has defective plumage. Frizzle crossbreds are characterised by curling of the feathers upwards and outwards. The pure-bred Frizzle remains practically bare throughout its first year of life, appearing to be in a state of perpetual moulting. It is extremely delicate and difficult to rear. When newly

hatched, the down feathering is fragile and breaks off. The exposure of the skin so produced leads to a great loss of bodily heat from the surface, calling forth increased basal metabolism, increased heat production, increased heart rate, lack of fat deposits, and diminished haemoglobin content of the blood. American biologists have now studied the genetic physiology of this breed, and have shown that the pure Frizzle chick will develop within three weeks a complete plumage over the whole body if protected from heat loss by enclosure in a woollen jacket and confined to a warm room.

Thus knowledge of the way in which a single dominant gene substitution produces its deleterious manifestations teaches us how to prevent their appearance. Researches of this kind have emphasised two important conclusions. One is the need for defining the kind of environment in which a given gene substitution manifests itself in a particular way. A second is that one and the same gene substitution may be responsible for many and various manifestations, depending upon the kind of environment in which development occurs. In the fruit fly *Drosophila* one gene is predominantly effective in the production of eye colour but has an accessory effect upon the wings. Another influences the number of bristles but has measurable effects upon at least a dozen other characteristics to a less noticeable extent. For convenience we usually define a gene substitution by the most striking effect which it produces in some specified environment or by the single effect which it produces in the widest range of environment in which its effects can be recognised. In reality no gene can be supposed to have a single absolutely specific effect.

The effect of a gene substitution depends on all the other genes with which it is combined. An example from the pathology of fishes will illustrate this. Two American biologists have recently made a study of intergeneric crosses between different varieties of two kinds of fish kept for ornament in aquaria. Their popular name is the Mexican killifish. Varieties of *Platypoecilus* differ in possessing large black pigment cells, small black pigment cells, or no black pigment cells at all. In inter-specific crosses, the occurrence of the two kinds of black pigment cells can be shown to depend respectively upon a sex-linked and an autosomal dominant gene substitution. Crosses between genus Xiphophorus and varieties of Platypoecilus having large black pigment cells result in the production of offspring with tumours. Thus a gene substitution whose effect is merely ornamental

and, as such, of commercial value, when accompanied by one combination of genes, is definitely pathological in its effect in the presence of another.

If two stocks are cultured side by side in the same environment so that any relevant idiosyncrasies of nurture are distributed equally between individuals belonging to each stock, we can tell that they are genetically different when a consistent average difference in the measurement of any of their characteristics reappears in successive generations. This is the most general definition of a genetic difference, and includes as a special case varieties of which any individual member differs from its group mean by an amount less than would any individual of another variety. The classical researches of Mendel with tall and dwarf peas were not made possible because all individuals of the tall strain are exactly the same size in any environment capable of sustaining them. It would be quite easy to arrange conditions in which they would be stunted and actually smaller than plants of the dwarf kind reared in the usual way. The fact is that if peas of a tall and dwarf strain are grown together with as much attention as any sensible gardener gives to his work, the shortest peas of the tall strain are recognisably taller than the tallest of the dwarf variety.

Differences of this kind—*unit characters* or discontinuous variations—do not exhaust the subject-matter of genetics. They occupied the attention of an earlier generation of experimental geneticists almost exclusively because they are easy to deal with, and they are easy to deal with because any individual of one stock can be readily distinguished from every individual of another when reasonable precautions are taken. If we cared to do so, we could express the difference between two stocks distinguished in this way by two average measurements. That would be unnecessarily sophisticated.

When we are dealing with gene differences which only manifest themselves within a very narrow range of environment, we may be able to recognise gene differences between two stocks in terms of such averages, even though we cannot decide whether an isolated individual belongs to one stock or the other. One of the most fruitful results of modern genetic analysis is the conclusion that a close system of inbreeding separates a mixed stock into genetically pure lines. This is implicit in the mathematical form of Mendel's principle, and has been abundantly proved to be true by such experimental work as that of Johannsen on beans and Helen Dean King

B

on rats. So, when the difference between two pure lines can only be expressed in terms of two average measurements for overlapping populations, it is reasonable to regard the individual variation in each stock as the effect of environment, unless the spread is of the same order of magnitude as the errors of measurement.

When we are studying animals in the laboratory we can arrange the conditions of an experiment so as to isolate gene differences or differences due to environment for separate treatment. We can use a highly inbred stock of rats to find how body weight varies with the vitamin content of the food or whether they form tumours when the skin is treated with pentacyclic hydrocarbons. If we keep all our rats on the same diet, we can also separate pure lines with different growth rates and greater or less resistance to tumours. With human populations the unaided investigator cannot do this sort of thing, and when we speak of heredity or environment as more or less *important* in connexion with any differences between human beings, *our criterion of importance is relative to the historic environment in which the differences themselves are measured.* Two hundred years ago the majority of Englishmen ran the risk of smallpox infection. No doubt gene differences played a large part in deciding whether a particular Englishman succumbed to the disease or escaped. No biologist or clinician would argue that gene differences provide the main reason why modern Englishmen are less likely to get smallpox than their great-grandfathers or than Esquimaux communities at the present day. We have created an environment in which it does not matter either way. In the course of millennia it is not unlikely that European communities could evolve a high degree of immunity to smallpox through uncontrolled selective elimination of the less resistant. The African peoples have probably evolved their high immunity to malaria in this way. Thanks to human inventiveness, we have not had to wait several millennia to get rid of smallpox.

Practical husbandry and scientific crop production provide various examples of how human valuations placed upon genetic differences are relative to the environment in which they are recognised. In his book *The Causes of Evolution*, Haldane cites two botanical illustrations. Engledow (1925) found that when two varieties of wheat known as Red Fife and Hybrid H are spaced at 2 inches by 2 inches, Red Fife yields the larger crop. At 2 by 6 inches the yields are almost equal, and at greater distances Hybrid H

yields a better crop than Red Fife. Sax (1926) has compared the crop of two colour varieties of the bean. Generally the white of his experiments yielded a smaller crop, but in exceptionally favourable conditions their yield was better than that of the coloured variety.

The recognition of a genetic difference thus implies one of two things: (*a*) That the difference is one which manifests itself in almost any environment suitable to the survival of the individuals concerned, or (*b*) that we càn reproduce the kind if environment in which it will be recognisable. The last statement is illustrated by the improvement of livestock in the eighteenth century. This happened because the introduction of root crops made it possible to standardise methods of feeding. Thus Ernle (*English Farming— Past and Present*) tells us:

> Bakewell's success and the rapidly increasing demand for butcher's meat raised up a host of imitators. Breeders everywhere followed his example; his standard of excellence was gradually recognised. The foundation of the Smithfield Club in 1798 did much to promote the improvement of livestock. Some idea of the effect produced may be gathered from the average weights of sheep and cattle sold at Smithfield Market in 1710 and in 1795. In 1710 the average weight for beeves was 370 lb., for calves 50 lb., for sheep 28 lb., for lambs 18 lb. In 1795 beeves had risen in average weight to 800 lb., calves to 148 lb., sheep to 80 lb., lambs to 50 lb. This enormous addition to the meat supply of the country was due partly to the efforts of agriculturists like Tull, Townshend, Bakewell, and others, partly to the enclosure of open fields and commons which their improvements encouraged. On open fields and commons, owing mainly to the scarcity of winter keep, the livestock was dwarfed in size and weight. Even if the number of animals which might be grazed on the commons was regulated by custom, the stint was often so large that the pasture could only carry the smallest animals. Where the grazing rights were unlimited, as seems to have been not unusually the case in the eighteenth century, the herbage was necessarily still more impoverished, and the size of the livestock more stunted. On enclosed land, on the other hand, the introduction of turnip and clover husbandry doubled the number and weight of the stock which the land would carry, and the early maturity of the improved breeds enabled farmers to fatten them more expeditiously.

Just as centuries of experience in mining, dyeing, and medicine were necessary to clarify the concept of a pure substance before theoretical chemistry could begin to flourish, centuries of experience in agriculture, stockbreeding, and horticulture preceded and contributed to the recognition of those so-called unit characters with which the pioneers of animal and plant genetics occupied themselves. Historically the recognition that certain characteristics regu-

larly reappear in certain stocks and do not do so in others went hand in hand with the practical task of designing the most favourable conditions for their appearance. The geneticist makes his appearance, when that task is accomplished.

§ 2

The stud-book method of Bakewell furnished the raw materials of the first and still the most successful discoveries about human inheritance. They were made by collecting family pedigrees of individuals with congenital deformities and diseases of the body. The data contained in pedigrees can yield valuable information, if the mutant genes responsible for an idiosyncrasy exert their effect throughout a wide range of environment. It is then possible to apply numerical tests to detect their presence; and a large list of physical conditions pass the tests satisfactorily.

In studying inheritance in human beings, it is not possible to start with pure-bred stocks. So if a human trait is recessive, that is to say, if it is only manifest when the individual receives a particular gene from both parents, a certain proportion of individuals who do not manifest the same trait receive the gene from one but not from the other parent. Similarly, if the trait is dominant, that is to say, if it is recognisable when the individual who shows it receives a particular gene from one parent only, it may not be possible to tell from the appearance of any given individual whether he or she has received it from one or both parents. Till recently this presented a considerable obstacle to the verification of quantitative laws in the study of human inheritance. It is no longer an insuperable difficulty. Marriage is a lottery. The natural history of lotteries, or, as we more usually call it, the theory of algebraic probability, enables us to predict what proportion of individuals will derive a given gene from both parents or from one parent only, if we know the proportion who do not possess it. Thus the net expectation for different kinds of offspring of parents of a specified type can easily be calculated if mating occurs at random. Mating does not always occur strictly at random in human communities. A talented contemporary authoress has pointed out that gentlemen prefer blondes. Allowance can be made for this by studying the correlation between husbands and wives. The next chapter will deal with the way in which the theory of random mating can be used in human genetics, and when it is not strictly applicable.

If mating occurs at random, it is easily shown that the number of individuals who carry a rare gene on one chromosome but not on its fellow is twice the square root of the number who carry it on both members of the same pair of chromosomes. What this means may be illustrated by albinism. Albinism is a recessive condition. In this country the proportion of albinos in the community is about one in twenty thousand. According to the principle of random mating, one in every seventy individuals who are not albinos should therefore carry the gene for albinism on one of their chromosomes. Thus individuals who display a very rare dominant condition will nearly always possess the gene which determines it on one chromosome only. Genetic theory demands that half the offspring of such individuals, if married to a normal person, will have the dominant trait. This is easy to test in the numerous pedigrees of what medical men refer to as' "hereditary" diseases or disfigurements. Such are brachydactyly, a congenital absence of one of the joints of the fingers, one form of night blindness, a somewhat repulsive abnormality known as lobster claw which is a deformity of the lower limb, the disease known as diabetes insipidus, Huntingdon's chorea, and the eye defect called aniridia. These conform in a satisfactory way to the numerical requirements of Mendel's law. They could be eliminated in a generation if individuals suffering from them were not allowed to reproduce. When diseases of this class are incurable, this is the only effective method of prevention known at present.

It might seem more difficult to identify recessive genes in human beings. An individual who exhibits a recessive condition must receive the gene from both parents. He or she may thus be the offspring of one of three types of marriage: a marriage between two recessives; a marriage between a recessive and an apparently normal individual who carries the gene; or a marriage between two carriers neither of whom exhibit the trait. What has been said about albinism shows that marriages of the last type will be vastly more common than the other two. In other words, recessives are generally offspring of parents who are not themselves recessives and have no near ancestors who are recessives. They will not be detected by collecting long pedigrees. We have to resort to other means.

Genetic theory tells us that if two parents are carriers one-quarter of their offspring will be recessives. Thus recessive conditions tend to turn up among several brothers and sisters in a family. In the language of the medical profession they are "familial." The pro-

portion predicted by genetic theory is easily tested by collecting
sufficient cases. In the fourth chapter we shall examine how the test
is applied. A second criterion which forms the subject of the third
chapter is still more valuable, especially if the recognition of a
recessive gene substitution depends on conditions which are not
always present in the family environment. Consanguineous parentage
will always be more common among parents of recessives than
among the general population. The proportion of consanguineous
parentage can be stated precisely as a function of the rarity of the
recessive condition. About 15 per cent. of the parents of children
who die of Tay Sachs disease are first-cousins. The percentage of
all marriages between first-cousins in the population at large
generally varies between one-half and 1 per cent. in European
communities. Without recourse to precise mathematical treatment,
the reason for this is easy to grasp, though unaided common sense
is not sufficient to tell us how rare a recessive condition must be
if we are to detect a large enough excess of consanguineous parentage.
If I carry the gene for albinism on one of my chromosomes, the
chance that I shall marry an unrelated individual who is likewise a
carrier is only one in seventy. If I marry my cousin, I am marrying
an individual who has received a certain proportion of her chromo-
somes from the same pair of grandparents as myself. The chance
that the offspring of two grandparents will both receive a particular
chromosome from one of them is one in eight. Hence, if I am
myself a carrier for albinism, the odds in favour of marrying another
carrier would be nearly ten times greater than if I married someone
who was not related to me.

About a dozen of these recessive conditions are now well estab-
lished. One is the familial type of retinitis pigmentosa described
by Usher. Amaurotic family idiocy and juvenile amaurotic idiocy
are two other examples. If two parents produce an amaurotic child,
the odds are that one-half of their offspring will carry the gene,
and one-quarter will exhibit it. It is difficult to justify the English
law which does not permit such parents to avail themselves of a
very simple operation to prevent the further spread. of the unwel-
come genes which are responsible for these two formidable and at
present quite incurable diseases. Sterilisation of the individuals
directly affected is in this case undertaken by nature. The affected
individuals die before they can propagate their kind. Selection
eliminates recessive conditions very slowly. If all albinos were

sterilised in every generation it would probably take many centuries to reduce the incidence of albinism to half its present dimensions.

Concentrated medical research upon the incidence of consanguineous parentage is likely to increase our knowledge of recessive genes in the human species considerably. At present a larger number of dominant than of recessive mutations are known to exist among human beings. This is contrary to what occurs in most wild animals. In nature dominant mutations seem to be rare. Probably there are more recessive than dominant mutations in Man. The apparent rarity of recessive mutations may be due to the fact that the method of detecting them has only been recently perfected. This is borne out by the existence of one special class of recessive genes not included in what has been said hitherto. Recessive genes borne upon the sex chromosomes are easily recognised by the fact that recessive females are much rarer than recessive males. Red-green colour blindness is a case of this type of inheritance. Colour-blind males are at least ten times as common as colour-blind females.

Recessive genes known to be located on the sex chromosomes are more numerous than all the recessive genes at present known to be located on the remaining twenty-three pairs of human chromosomes. No doubt this is because the peculiar type of inheritance to which they give rise attracted medical interest more than a century ago in connexion with the study of haemophilia. There is a strain of haemophilia in the Royal Houses of Europe. It has been said that the sterilisation of individuals who display recessive conditions of the ordinary type produces very little effect because the genes are principally transmitted by individuals who do not exhibit the recessive condition. This is not true of sex-linked recessive conditions. Since the male has only one sex chromosome, all males who carry the recessive gene exhibit the recessive trait, unless special conditions of environment are essential to its manifestation. The eye defect called megalocornea and one type of hereditary optic atrophy* (Leber's disease) belong to this class.

* The recognition of different genetic types within a single clinical category may provide a stimulus to more refined clinical classification. A suggestive illustration of this is provided by Leber's disease. One form of Leber's disease is determined by a recessive sex-linked gene. Another group of pedigrees points to the manifestation of two dominant genes which are located on other chromosomes. It happens that the mean age of onset of the disease in the two groups of pedigrees is different by several years. Thus genetic analysis leads to a differential diagnosis which has eluded traditional methods.

Sterilisation of all individuals displaying sex-linked recessive diseases halves the proportion of persons affected in every generation.

The tendency of traits to stick together in the same pedigree has made it possible to construct maps of the chromosomes in animals and plants. All the distinguishable genes of the fruit fly and the sweet pea can be assigned to their respective chromosomes and to a particular locus relative to other genes on the same chromosomes as themselves. Most of the genes whose manifest.effects are easy to distinguish in human beings are rare. It is therefore exceedingly unlikely that we should encounter two in the same pedigree. For this reason the possibility of constructing a chromosome map of the human species seemed quite fantastic ten years ago. To-day the prospects are very hopeful.* The possibility of doing so has emerged from the study of the blood groups. The four blood groups depend upon three genes, one group being recessive, one group depending on the presence of one dominant gene, a third on the presence of another dominant gene, and the fourth on the presence of both dominant genes. The two dominant genes have arisen by mutation from one and the same recessive gene. Both cannot be present on one and the same chromosome. Population studies on hundreds of thousands of individuals have shown that the different proportions of these groups in different communities correspond to the requirements of the theory of random mating with extraordinary fidelity. Parents and offspring of more than five thousand families have been systematically examined. The results are in close agreement with what would be predicted, if the explanation already given is the correct one. Their importance for human genetics resides in the fact that the frequency with which the three blood-group genes occur in the general population is much the same. Hence it is easy to test whether they tend to stick together with other genes.

If blood-group testing were carried out in all records of clinical pedigrees, it would be possible to ascertain whether rare genes responsible for diseases like amaurotic idiocy or night blindness reside on the same chromosomes as the three genes of the blood groups. Agglutination can also be produced by injecting sera of other animals into the circulation. People have now been classified

* Bernstein (1931), "Zur Grundlegung der Chromosomentheorie der Vererbung bein Menschen," *Zeit. f. Indukt. Abstamm. und Vererblehre*, LVII. See Appendix VII for his method.

for their reactions to various foreign sera. Other blood groupings of similar proportions have been based on such reactions, and the transmission of at least one such series has been clearly established.* It has been shown that the genes involved are not located on the same pair of chromosomes as the three genes of the Jansky blood groups. There is a very hopeful prospect that we shall soon be able to test for a blood-grouping referable to every one of the twenty-four pairs of human chromosomes. Recently it has been shown that about a quarter of the population are incapable of tasting a group of substances allied to the organic compound called phenyl-thiourea. This substance is exceedingly bitter to those who can taste it. Ability to taste is determined by a single dominant gene. About as many people have the dominant gene as lack it. Like the blood-group test, this reaction may play a part in the mapping of the human chromosomes.

About thirty known incurable diseases are determined by genes whose existence is established by quantitative agreement with the requirements of Mendel's laws. This list includes several forms of blindness. Large-scale investigations such as those which are being carried out in Russia by Levit and Serebrovsky will probably extend the list very considerably in the near future. We do not yet know of any enviable characteristics of human beings determined by single genes. Even the inheritance of the platinum blonde is still obscure. The next few years may witness substantial progress in establishing precise laws of hereditary transmission for physical traits which are little affected by differences of environment to which members of the same fraternity are ordinarily subject.

§ 3

Several defects of neuromuscular organisation, such as *Amaurotic Family Idiocy*, all of them associated with detectable physical symptoms, are included in the list of established diseases which depend on a gene difference which is recognisable in all customary conditions of development. When the geneticist is confronted with a discontinuous character which only manifests itself in special circumstances, his first line of attack is to find out everything he can about how nurture controls its appearance.

* Landsteiner and Levine (1931), "The Differentiation of a Type of Human Blood by means of Normal Animal Serum," *Journ. Immunol.*, Vol. XX.

There is a mutant of the fruit fly which is deformed in the hindmost part of the body. It is usually referred to as "abnormal abdomen." Flies of the pure mutant stock regularly exhibit the deformity when grown in moist cultures. They are perfectly normal when grown in a dried-up culture. So long as the experiments are carried out in moist cultures, matings with wild-type stock yield numerical ratios in keeping with the supposition that the difference between the mutant stock and the wild type is due to a single gene substitution. If the cultures are allowed to dry, no consistent numerical results can be obtained for an obvious reason. Had the geneticist no means of preventing his cultures from drying up, he would have to confine his counts to larvae which hatch out while the culture is still moist. A type of feeble mindedness known as mongolism provides an example of how medical science can apply a similar method. Mongols rarely have young mothers. A high percentage of mongols have mothers about forty years of age or more. Since the genetic constitution of the mother is not affected by her age, the environment of the womb must have something to do with whether an individual is a mongol. By studying families born after the mother has passed a certain age, Dr. Penrose has been able to show what part heredity plays in producing mongolism.

Mongolian idiocy is associated with well-defined physical characteristics, to some of which it owes its name. So far the comparison of pedigrees containing feeble-minded or mentally deranged individuals with no distinctive physical stigmata has not yielded information which satisfies any numerical criteria based on genetic theory. This may be because the occurrence of imbecility and lunacy is due to a combination of genes too complex to detect without recourse to experimental mating which is impracticable. It may be because a particular combination of genes and a particular kind of environment are jointly responsible for producing them.

Of its very nature social behaviour depends on an environment complex which cannot be standardised. Individual differences of social behaviour, as we observe them, are differences to which differences of environment and gene differences jointly contribute. When differences of environment and differences of gene equipment jointly contribute to observed differences between human beings it may be that the responsible genes are rare (or are mainly confined to a small group of people), while the conditions of nurture on which their detection depends are relatively common. Heredity is then the

more significant source of variation. Conversely, it may be that the responsible genes are widely distributed in the population, while the relevant conditions of nurture are rare or very unevenly distributed. The more important source of variation, then, resides in the environment. In this sense we are entitled to ask whether nature or nurture is the most important agency which determines individual differences. The question can be investigated on a statistical scale when it is not possible to find out which decides the fate of a particular individual. Of several methods which can be used the three most important ones are (a) the method of twin resemblance, (b) the method of adoption, (c) the method of consanguinity.

The method of twin resemblance was first suggested by Galton. Partly because the pertinent facts were not fully established and partly because there were insufficient endowments to support large-scale research, it has not been applied extensively till recent years. Embryological research has shown that when mammals produce several offspring together the same result may be produced in different ways. Most species have litters of several offspring because several egg cells are set free into the womb when the mother is on heat. A few multiparous species liberate only one egg at a time. The mass of cells produced from the fertilised egg then splits at an early stage of development to form several embryos. When multiparous pregnancies occur in human beings and in cattle either process may be responsible. Hence human twins are of two kinds. *Identical* twins, being descended from the same fertilised egg, have the same set of genes and are necessarily of the same sex. *Fraternal* twins, being descended from different eggs, have sets of genes which are no more alike than those of ordinary offspring of the same parent. Such twins may be of like sex or unlike sex. They can now be distinguished from the other type by reliable tests.

This fact may be used to investigate the relative importance of nature and nurture in two ways. We may compare the degrees of similarity shown by identical twins, fraternal twins, and ordinary "sibs" (brothers or sisters) brought up together in the same family, and we may compare the resemblance of identical twins reared apart with that of identical twins brought up together. If identical twins are decidedly more alike than fraternal twins in the same family, we may conclude that heredity plays a large part in deciding the difference between individual members *of a single family*. If fraternal twins are decidedly more alike than ordinary sibs we may

conclude that the differences of environment to which children of different ages, brought up in the same family, are exposed, play a large part in deciding the characteristics of individual members. Since the environment of a family at one social level may be very different from that of a family at another, the discovery that heredity is the chief agency which decides what the characteristics of different members of the same fraternity will be, does not necessarily imply that it is the chief agency which decides differences between individuals belonging to different social classes, races, or religions.

This can be settled by comparing the degrees of similarity shown by identical twins reared together and identical twins reared apart in totally different social circumstances. It happens when they are adopted at birth, because their parents die or desert them. The practice of adoption can also be used in another way. If true sibs reared together are decidedly more alike than true sibs reared apart, or if foster sibs are more alike than pairs of individuals taken at random from similar homes, differences of home environment may be inferred to play a decisive rôle.

A third method of investigating the rôle of nature and nurture depends on the theory of inbreeding. Inbreeding results in separating *pure* stocks from a hybrid population. Hence it increases the amount of variety. It is not difficult to see that this is true where the number of genes involved is small. The reason for this is the same as the reason for the high proportion of albinos whose parents are first cousins. Hence a high measure of variability among children whose parents are consanguineous when compared with children whose parents are not related points to the influence of nature rather than of nurture.

These methods of attack have been elaborated within the last twenty years. That they have been applied to the study of comparatively few aspects of man's social behaviour is chiefly due to two circumstances. The first is the persistence of the stud-book mentality.* The overwhelming majority of publications ostensibly dealing with human heredity are collections of pedigrees. The analysis of pedigrees can supply useful information when the data supplied by them satisfy numerical tests suggested by the known behaviour of genes. The fact that they pass the tests justifies the suggestion that ordinary differences of environment do not interfere with the

* Cf. the Lidbetter pedigrees and Dr. Hurst's memoirs on the inheritance of intellectual ability.

expression of the gene difference. So the conclusions drawn from them are irresistible. When the data supplied by pedigrees fail to do so we are in doubtful territory, and the more so when we are studying social characteristics such as temperamental traits and intellectual performance, which are known to demand certain limiting circumstances of upbringing. The stud book is a reliable guide to the inborn qualities of pedigree cattle, because the farmer aims at equalising the environment of individuals selected for parenthood. For two reasons it is not a reliable guide to the contribution which heredity makes to differences of behaviour. One is that the human family transmits a certain social tradition, i.e. a particular sort of environment as well as a certain equipment of genes. One is that equality of environment is not yet the recognised goal of social organisation, least of all by most eugenists. The studbook method is used because those who profess to accept the stud farm as a model for human betterment shrink from promoting the social arrangements which would make the analogy pertinent to the circumstances of social life.

Another serious obstacle to progress is the paucity of methods for measuring and describing differences of social behaviour. A beginning has been made with the intelligence tests of Binet, Terman, Burt, Spearman, and others. When people apply the word intelligent to a person they do not mean something as definite as black, freckled, or intoxicated. This does not imply that no useful meaning can be attached to the word *intelligent* as a description of the characteristics of human beings. Different observers can arrange a group of individuals in a scale of what they call greater or less intelligence. They can then see whether their arrangements tally and whether it is possible to devise some independent test by which the same group can be arranged in a way which corresponds fairly closely with independent estimates based on personal impressions. This is what an intelligence test does. Extensive and careful statistical researches have been undertaken to devise a scale which will record what is *common* to the various ways in which people use the word *intelligent*, when they apply it to the social behaviour of children and adolescents. It does not necessarily follow that the intelligence tests give a just measure of all that we commonly mean by the adjective intelligent when we apply it to adults. Probably the intellectual performance of adults depends quite as much on temperamental characteristics ordinarily described by alertness, persistence,

curiosity, or a sense of humour as on the type of facility which intelligence tests assess. Hence proposals to limit educational facilities to children who get high scores in such tests are exceedingly dangerous. It is never suggested that the education of the prosperous classes should be limited in the same way. So the political motive is not far to seek.

The great advantage of the tests on which such scales are based is that they yield very constant results for the same individual examined on successive occasions if the intervening period is short. They also give fairly constant results for the order of individuals within a group when it is tested successively over a period of several years. Dubious speculations sometimes built upon this solid foundation of fact need not concern the geneticist or the clinician. What is important for our purpose is that we now have a method of describing one aspect of human behaviour with some precision and reliability. It can be passed from the hands of one observer to another. So we can pool the results of intelligence tests as we could not do if we had to rely on any customary scale such as teachers' estimates, examination results, or employers' testimonials.

This means that the biologist can investigate to what extent differences of intelligence are associated with the fact that different children are born with different genes, and how far the manifestation of such gene differences is independent of maternal health in prenatal existence, other conditions of uterine environment, a poorly nourished body, over-indulgent parents, overbearing brothers and sisters, sympathetic teachers, and an infinite variety of other circumstances which distinguish the physical and social environment of one individual from another. On the evidence at present available, it seems the average differences between the intelligence quotients of members of the same fraternity is reduced by less than half when all gene differences which differentiate offspring of the same parents from one another are eliminated. There is only one piece of unequivocal evidence to show that heredity makes *any* contribution to individual variation in the middle of the normal range of intelligence test scores. Identical twins reared apart are difficult to find, and the social machinery of adoption usually places them in homes *at the same social level*. The fragmentary evidence available show that the average IQ difference of 10 pairs of identical twins reared apart is 7·7 points. The average IQ difference for fraternal twins reared together is only 9·9 points. If we had a large sample of

identical twins reared apart in homes at different social levels the difference might well be greater than 7·7. Till then the reader may judge for himself whether scare headlines about the decline of the nation's intelligence belong to the province of science.

This does not exclude the possibility that future research may detect and measure racial differences of intelligence depending upon differences of genetic constitution. The difficulty of treating group differences of this kind in a genuinely scientific temper will be less when psychology can equip biological research with a sufficient variety of similar methods for the precise description of other aspects of social behaviour. One can assert that deaf-mutism is commoner among Jews than among Gentiles without incurring the charge of anti-Semitism. With so many diagnosable physical ailments to choose from, it is possible for normal people to discuss the occupational or racial distribution of any single disease of the body without assuming a tone of impudent superiority. No single group' has the monopoly of all the virtues. Time may show that there are genes which have something to do with many estimable attributes. We shall then see the value of superior intelligence in a proper social perspective.

§ 4

That biologists do not always give the same answer to questions about heredity and disease is partly due to the fact that questions framed in everyday speech involve an ambiguity which arises from the changing nature of the human environment. If we ask, is amaurotic idiocy associated with a gene difference which manifests its presence throughout the whole range of conditions to which members of the same fraternity are normally exposed, we can expect a biological answer. No biologist who is conversant with the facts will hesitate to answer the question in the affirmative. Few sensible people who know the answer would encourage a married couple who had produced a child with the disease to have more children. If we ask, Is simple primary amentia inherited? many biologists will answer in the affirmative. They will not do so because we possess precise knowledge about the gene differences which affect feeble-mindedness of the same definite and unequivocal kind as our knowledge about amaurotic idiocy. They will interpret the question in a sociological sense, and give it what is implicitly a sociological answer. The answer conceals the assumption that if we

cannot control the environment we ought to take no chances with the hereditary aspect of the problem. This is not unreasonable so long as there is no disposition to underrate the likelihood of discovering how to control the environment.

Of late years the discussion of preventive applications of genetic theory has been largely focussed upon the problem of mental defect. There is no doubt about the concentration of mental defect of one kind or another in certain inbred stocks, and such concentration is not likely to be due entirely to the family environment. If it were, most members of a fraternity would be alike. On the other hand, the fact that the family is a unit of social and physical environment, and that the ancestry of a human being is a complex of genealogical and environmental relationships, is entirely consistent with the view that genetic differences manifest in such pedigrees would not necessarily manifest themselves in other situations.

It is clearly desirable to study the genetic aspect of feeble-mindedness and to make the best of all the knowledge we gain. At the same time there is no occasion to magnify its importance till it assumes menacing dimensions. It can be maintained with some reason that it would be wiser to prevent the feeble-minded from propagating in the meantime, because they are obviously unsuitable parents from an educational point of view. A rationally planned society might easily be persuaded to take this course. The sympathy with which such proposals are at present greeted by some students of social problems is not enhanced by the fact that they are almost invariably put forward by those who are most anxious to perpetuate forms of parasitism more costly and disastrous than feeblemindedness. To the writer it seems that the selfishness, apathy and prejudice which prevent intellectually gifted people from understanding the character of the present crisis in civilisation is a far greater menace to the survival of culture than the prevalence of mental defect in the technical sense of the term.

This does not mean that the study of human inheritance is unimportant. On the contrary it has everything to gain by outgrowing the castration complex. With the prospect of a spectacular decline of population in the near future constructive statesmanship will be more and more preoccupied with ways and means to encourage parenthood. Consequently it will be less and less favourable to drastic proposals for sterilising the harmlessly unfit. For the same reasons it will be more and more committed to an active

policy of preventive medicine. As part of an active policy of preventive medicine the future of human genetics is assured. No community is likely to sterilise people who suffer from frontal sinus infections, or to subsidise research which leads to the conclusion that people who suffer from sinus infections should necessarily be sterilised. What makes it important to know everything which can be found out about the contribution of heredity to such diseases is that if we have such knowledge we can forewarn people who are liable to contract them against exposing themselves to the dangers of infection. So long as sterilisation is the goal of human genetics, its scope must be. limited to the study of comparatively serious disorders. As a department of preventive medicine it embraces the whole field of disease.

Analogous remarks apply to education. Eugenists are never tired of talking about the "waste" of expenditure on those who are "by nature" unable to benefit from it. Naturally this does not engage the sympathy of educationists who take their job seriously. Nor does it enlist the support of intelligent citizens, who realise that no society is safe in the hands of a few clever people. If knowledge is the keystone of intelligent citizenship, the fact that many people do not benefit from existing provisions for instruction is less a criticism of themselves than a criticism of educational machinery. The possibility that heredity plays a large part in such differences is only relevant to public expenditure, when we have already decided whether we want more or less education. We do not need biologists to tell us that any subject can be made dull enough to defy the efforts of any but a few exceptionally bright or odd individuals. By exploring individual differences human genetics might help us to find out how to adapt our educational technique to individual needs. It will do so, and gain prestige in consequence, when it ceases to be an apology for snobbery, selfishness, and class arrogance.

APPLICATIONS AND LIMITATIONS OF THE PRINCIPLE OF RANDOM MATING

§ 1

The quantitative data of modern genetics are generally interpretable on two simple assumptions. The first is that differences which exhibit themselves in a homogeneous environment in virtue of the ancestry of the individuals which show them are due to differences in material units (*genes*) with a definite position upon the chromosomes. The second is that the distribution of the constituent members of a chromosome pair in the process of reduction and the pairing of gametes containing different genes takes place at random. On this hypothesis new genetic or *mutants* arise by the substitution of one or more genes for those already present in the chromosomes. The application of these principles to the results of matings between two pure stocks differing with respect to a single gene substitution lead to two general conclusions which are true for all genes except those which are located upon the X-chromosomes. The first is that crossbreds from pure parents if mated *inter se* produce offspring of which one-half are like their crossbred parents and one-quarter respectively like each of their pure-bred grandparents in the genes they possess. The other is that matings from crossbreds and either type of pure parents produce offspring half of which resemble the pure parent and half the crossbred parent in the genes they possess.

There would be no difficulty in applying these generalisations to human data if the effects of gene differences were simply additive. In that case hybrids would always be intermediate between their two parents like the blue hybrid of black and white Andalusian fowls. Crossbreds are generally more like one parent than the other. Often they cannot be distinguished individually from the parent they resemble most, though statistical methods show that they are never quite the same. Thus mutants tend to belong to one of two extreme types: *dominant* mutants arising from a gene substitution whose presence is manifest when present on one chromosome only, and *recessive* mutants which are only recognised as such when the sub-

stituted gene is present on both members of a pair of chromosomes. In fowls the "rose" comb of the Wyandottes may be attributed to a dominant gene substitution which occurred in the ancestral stock of fowls with "single" combs.

The animal geneticist starts with pure stocks. Throughout the multitudinous marriages which he solemnises he records the precise ancestry of each individual. He can test the results of mating two pure rose-comb or two pure single-comb birds, pure rose comb or crossbred rose comb with single comb, pure rose comb with crossbred rose comb or two crossbred rose-comb birds with one another. The human fancier, the man with a fancy for fitting genetic hypotheses to family histories, has a different problem. He can record what happens in the human poultry run when two single comb, two rose comb, or a rose-comb and a single-comb bird mate together. He does not know whether any particular rose-comb bird is pure or crossbred. The fact that marriages between two single-comb birds never produce rose-comb offspring may suggest that the single comb is the recessive condition. Further than that he cannot go unless he makes certain assumptions about how mating takes place.

The truth of a genetic hypothesis rests upon quantitative agreement of the results predicted and the phenomena observed. To make such predictions we must either be able to distinguish the several genotypes by their appearance and ancestry or else possess some information concerning the relative frequency with which matings of different genotypes occur. In considering the behaviour of different pairs of chromosomes and of gametes with different genes the animal geneticist makes calculations based on the analogy of drawing coloured balls from a bag. His physical model is the lottery. The human geneticist who cannot distinguish the pure-bred from the crossbred dominant genotypes can fall back on the same model. He can then predict the proportions in which different types of mating will occur, and test whether it is possible to find an hypothesis consistent with the assumption that mating occurs at random.

At a later stage we shall illustrate how this can be done by examples from the study of human inheritance such as the phenomenon of *taste blindness*. This is the inability to recognise the taste of a certain group of organic compounds. The inheritance of taste blindness in human beings is precisely analogous to the inheritance of the rose and single comb in the fowl. Marriages between two people of

the class described as "non-tasters" only produce offspring of the same class, just as two single-comb fowls always have offspring with single combs. Marriages between two tasters or between a taster and a non-taster may produce offspring of either type, just as matings between two rose-comb fowls may produce either rose-comb or single-comb chicks if the rose-comb parent is heterozygous. Tasters like rose-comb fowls may be crossbred or pure bred in the genetic sense. We cannot distinguish between individuals of one or the other kind. The Principle of Random Mating allows us to deduce in what proportions tasters and non-tasters will occur among off-spring of marriages between two tasters or marriages between a taster and a non-taster. First it is necessary to establish the Principle of Random Mating. Then we shall examine when it can be used successfully. Finally we must examine when it breaks down. A scientific law is only correctly stated when it contains within itself a recognition of its own limitations.

The hypothesis that mating occurs at random is the simplest one to adopt. Though we may be disposed to think that the results could not be more disastrous if men selected their partners by taking a name at random from a telephone directory, we should not apply the theory of random mating without the closest scrutiny. In certain circumstances it is obviously untrue. A young woman with a strong belief in the sprinkling of babies may be inclined to reject the advances of a young man with an equally strong conviction in favour of the total immersion of believers. Even when we are dealing with differences which do not obviously affect the choice of partners, it is necessary to remember that random mating in the strictly mathematical sense implies that every individual has equal access to every other individual in an infinite population. Human popu-lations are not infinite. In small rural populations inbreeding may give rise to considerable departures from a treatment based on strictly random mating. The effective population from which a diffident human being can choose a mate is always restricted and often quite small. Different genetic types are rarely distributed with perfect uniformity throughout a population. For all these reasons random mating is at best an approximation which should only be applied to large populations with great circumspection.

These objections do not dispose of the possibility that the theory of random mating can sometimes provide a close approximation to the distribution of genes in a human population, and may be used

as a basis for genetic analysis subject to further refinement arising from research upon the extent to which assortative mating occurs. Researches upon assortative mating have been carried out by Pearson and his colleagues, who have determined correlations of husband and wife for such characteristics as stature and eye colour. Such correlations merit very careful interpretation and very cautious application because of the lack of uniformity with respect to the distribution of genes in a community. If married couples are taken from a very large geographical area high correlations may be obtained, because of the concentration of individuals with similar traits in different localities. Much lower correlations might be obtained if the same number of married couples were taken from a single locality. Analogous considerations apply to the theory of random mating. Major brachydactyly in Great Britain has a frequency of about one in a million. Most brachydactyls are confined to a small locality in North Wales. It would be absurd to use the frequency of brachydactyly in the entire population to calculate the frequency of matings between brachydactyls and normal persons.

§ 2

When mating occurs at random in a population of mixed genetic types the proportions of the several genetic types reach equilibrium. Afterwards they do not change from one generation to another in the absence of selection. When only one gene substitution is involved these proportions can be stated definitely, if we know the frequency of any one of them. In what follows the symbols R, H, D will be used for the Recessive, crossbred Dominant (Heterozygote) and pure Dominant respectively. If we suspect that a condition is recessive, we can find out the proportion of supposedly recessive individuals in the country as a whole. The Principle of Random Mating then tells us the relative frequencies of the two dominant types (H and D), the relative frequencies of the two matings between a recessive and a dominant ($R \times H$ or $R \times D$) and the relative frequencies of matings between two dominants ($D \times D$, $D \times H$ or $H \times H$). As applied to single gene substitutions which do not involve the X-chromosome the Principal of Random Mating may be stated in two propositions:

(i) A population is in equilibrium when the proportion of

heterozygotes is twice the square root of the product of the frequencies of the two homozygous classes.

(ii) Equilibrium is attained in a single generation of mating at random, so that if anything occurs to displace the pre-existing equilibrium a new equilibrium is reached after mating has once occurred.

(1) Consider a population R_n H_n D_n in the proportions $\alpha:\beta:\gamma$ which may have any values consistent with the relation $(\alpha + \beta + \gamma) = 1$. The possible matings may be set forth in the familiar chess-board schema as follows:

	αR	βH	γD
αR	α^2	$\alpha\beta$	$\alpha\gamma$
βH	$\alpha\beta$	β^2	$\beta\gamma$
γD	$\alpha\gamma$	$\beta\gamma$	γ^2

Recessives may arise from matings $R \times R$ of which there are α^2, $R \times H$ of which there are $2\alpha\beta$ and $H \times H$ of which there are β^2. In the first case all offspring, in the second half and in the third one-quarter are recessive, so that in the ensuing generation the recessives are

$$R_{n+1} = \alpha^2 + \alpha\beta + \tfrac{1}{4}\beta^2$$

Similarly

$$H_{n+1} = 2\alpha\gamma + \alpha\beta + \beta\gamma + \tfrac{1}{2}\beta^2 = 2\alpha\gamma + \beta - \frac{\beta^2}{2}$$

$$D_{n+1} = \gamma^2 + \beta\gamma + \tfrac{1}{4}\beta^2$$

Now if the population is in equilibrium the proportion of any class in two consecutive generations is the same.

i.e. $R_n = R_{n+1}$, $H_n = H_{n+1}$: $D_n = D_{n \times 1}$

Hence we may put

$$\alpha = \alpha^2 + \alpha\beta + \tfrac{1}{4}\beta^2$$
$$= \alpha^2 + \alpha(1 - \alpha - \gamma) + \frac{\beta^2}{4}$$

Therefore

$$\beta^2 = 4\alpha\gamma$$

or

$$\beta = 2\sqrt{\alpha\gamma}$$

Putting $a = a^2$ and $y = b^2$ it is more convenient to represent the equilibrium state in the following form:

R	H	D
a^2	$2ab$	b^2

Since $a^2 + 2ab + b^2 = 1 = (a + b)^2$, we may write $(a + b) = 1$, so that the proportions of R, H and D in the equilibrium state are

$$a^2 : 2a(1 - a) : (1 - a)^2$$

Here a and b have a simple genetic significance. If the fraction a is the probability that the recessive gene will be borne upon a given chromosome, the probability that it will be present on that chromosome and also on its fellow is a^2, and the probability that it will be on either member of a given pair of chromosomes but not on both simultaneously is $2a(1 - a)$. Hence we may speak of a and b as the gene frequencies of the recessive and dominant allelomorphs. With the aid of the chess-board schema the reader will at once see that the frequencies of the various kinds of matings are:

$$a R^4 R : 4a^3 b R H : 2a^2 b^2 R D : 4a^2 b^2 H H : 4ab^3 H D : b^4 D D$$

(2) To establish the second proposition we must retrace our steps to the original statement of the problem when α, β, γ were given any values consistent with the relation $(\alpha + \beta + \gamma) = 1$. That is to say, we do not assume that the nth generation has the same consistution as the $(n + 1)$th, or that it has been produced by random mating. Then if the $(n + 1)$th generation is produced by random mating we have:

	R	H	D
nth generation	α	β	γ
$(n+1)$th generation	$\alpha^2 + \alpha\beta + \dfrac{\beta^2}{4}$	$2\alpha\gamma + \beta - \dfrac{\beta^2}{2}$	$\gamma^2 + \beta\gamma + \dfrac{\beta^2}{4}$

If equilibrium is attained in a single generation of mating at random,

$$R_{n+1} = R_{n+2}$$

It is easily seen that this is so. The form of R_{n+2} can be written down from the above, viz.:

$$R_{n+2} = \left(a^2 + a\beta + \frac{\beta^2}{4}\right)^2 + \left(a^2 + a\beta + \frac{\beta^2}{4}\right)\left(2a\gamma + \beta - \frac{\beta^2}{2}\right)$$

$$+ \tfrac{1}{4}\left(2a\gamma + \beta - \frac{\beta^2}{2}\right)^2$$

$$= a^2 + a\beta + \frac{\beta^2}{4} = R_{n+1}$$

(3) The distribution of sex-linked genes in a system of random mating obeys a different law. If $(a + b) = 1$, a population is in equilibrium when the proportions are as follows:

	Females			Males	
R	H	D		R	D
$\dfrac{a^2}{2}$	ab	$\dfrac{b^2}{2}$		$\dfrac{a}{2}$	$\dfrac{b}{2}$

As before, a and b are the gene frequencies of the recessive and dominant allelomorphs for the population as a whole. That a population with this constitution is in equilibrium is seen thus. As before, the frequencies of all possible types of matings is obtained by the aid of a chess-board schema.

		Females		
		R	H	D
		$\dfrac{a^2}{2}$	ab	$\dfrac{b^2}{2}$
Males	$\dfrac{a}{2}$ R	$\dfrac{a^3}{4}$	$\dfrac{a^2b}{2}$	$\dfrac{ab^2}{4}$
	$\dfrac{b}{2}$ D	$\dfrac{a^2b}{4}$	$\dfrac{ab^2}{2}$	$\dfrac{b^3}{4}$

Thus the frequencies of the several matings are in the proportions

$RY \times RR$	a^3	$DY \times RR$	a^2b
$RY \times RD$	$2a^2b$	$DY \times RD$	$2ab^2$
$RY \times DD$	ab^2	$DY \times DD$	b^3

The frequencies of the ensuing offspring may be represented in tabular form thus:

Mating	RR	RD	DD	RY	DY
$RY \times RR$	$\frac{1}{2}a^3$	—	—	$\frac{1}{2}a^3$	—
$RY \times RD$	$\frac{1}{2}a^2b$	$\frac{1}{2}a^2b$	—	$\frac{1}{2}a^2b$	$\frac{1}{2}a^2b$
$RY \times DD$	—	$\frac{1}{2}ab^2$	—	—	$\frac{1}{2}ab^2$
$DY \times RR$	—	$\frac{1}{2}a^2b$	—	$\frac{1}{2}a^2b$	—
$DY \times RD$	—	$\frac{1}{2}ab^2$	$\frac{1}{2}ab^2$	$\frac{1}{2}ab^2$	$\frac{1}{2}ab^2$
$DY \times DD$	—	—	$\frac{1}{2}b^3$	—	$\frac{1}{2}b^3$

Adding the columns, we find

$$RR = \frac{1}{2}(a^3 + a^2b) = \frac{1}{2}a^2(a + b) = \frac{1}{2}a^2, \text{ etc.}$$

Thus we have as before:

RR	RD	DD	RY	DY
$\frac{1}{2}a^2$	ab	$\frac{1}{2}b^2$	$\frac{1}{2}a$	$\frac{1}{2}b$

This condition of equilibrium is not attained in a single generation as with autosomal gene substitutions. If the pre-existing equilibrium is displaced, the new one is approached gradually, and the proportions in ensuing generations oscillate about their new equilibrium values.

The sex proportions of the several genotypes in a system of random mating draw attention to two conclusions which are of great assistance in the identification of sex-linked conditions. The ratio of recessive males to recessive females is $\frac{1}{2}a : \frac{1}{2}a^2$, or $1 : a$. In other words, affected females are far less common than affected males when a disease or deformity depends on a rare sex-linked gene. It is doubtful whether any authentic cases of haemophiliac women occur in the published pedigrees containing about 1,000 male haemophiliacs. If the incidence of the disease is taken as $1 : 100,000$, affected males should be 50,000 times as common as affected females. Hence the absence of well-authenticated female cases does not constitute any cogent argument in favour of the view that the recessive gene is unable to manifest its presence in the female soma. The ratio of dominant males to dominant females is

$$\frac{1}{2}(1 - a) : \frac{2a(1 - a) + (1 - a)^2}{2}$$

or

$$1 : 1 + a$$

If the dominant condition is very rare, a will not differ greatly from unity. So the ratio approximates to $1 : 2$. Thus affected females are roughly twice as common as affected males when a condition is determined by a rare dominant sex-linked gene substitution. As a practical illustration of the Principle of Random Mating applied to sex-linked conditions, the frequency of red-green colour blindness may be taken. Dr. Julia Bell gives the percentage of known males among recorded cases as $89 \cdot 2$. Thus colour-blind males are nine times as common as colour-blind females. In Europe, from which most of the data are taken, the incidence of colour-blind males is about 4 per cent. of all males, or 2 per cent. of the whole population. If the theory of random mating were strictly applicable, $a = 0 \cdot 04$, and males would be twenty-five times as common as females. The Chinese data, based on 1,200 patients examined by Field, give a better correspondence. The proportion of affected persons among males is $3 \cdot 2$, and males are nearly twenty times as common as females.

<div align="center">§ 3</div>

Two examples will now be given to illustrate the way in which these conclusions can be used to make verifiable predictions. Both deal with differences which are not likely to influence the choice of a mate. The first is the inheritance of what is called taste blindness, that is to say, inability to distinguish the taste of a group of synthetic compounds including phenyl-thio-urea and related substances. The second is the inheritance of the isoagglutinin reaction of human blood. Each involves characteristics which can only be recognised by scientific tests. For this reason they are not likely to influence the love affairs of people who are not chemists or biologists. The Principle of Random Mating may therefore be applied to them with some confidence of obtaining a satisfactory agreement between hypothesis and fact.

Taste blindness was discovered in 1931 by Fox. To many people —about two-thirds of an elementary school population in London— phenyl-thio-urea has a very bitter taste in very low dilutions. Others cannot taste it at all. Though individuals vary with respect to the threshold, they fall into two groups. The frequency distribution is sharply dimodal. So a population may be divided into "tasters" and "non-tasters" with very few doubtful cases. The genetic basis

of this difference was first investigated by Blakeslee and by Snyder (1931).* Both agree that if two parents are both "non-tasters" they only have offspring who are "non-tasters." The exceptions are so small in number that we can attribute them to doubtful paternity. This suggests that inability to taste is a recessive condition, and that ability to taste depends upon a single dominant gene subject to considerable somatic variability, or to a graded series of dominant genes each of which is sufficient to produce a recognisable sensitivity to this class of compound. Taking the simpler hypothesis, let us see what it implies if mating does occur at random.

When a difference between individuals involves a single gene substitution, only three kinds of matings can produce recessive offspring. We may tabulate them as follows:

Types of Mating	Proportion of Recessive Offspring
$R \times R$	1
$R \times H$	$\frac{1}{2}$
$H \times H$	$\frac{1}{4}$
$R \times D$ or $H \times D$ or $D \times D$	0

The hypothesis we have chosen is that non-tasters are all of the genotype R like single-comb fowls, and tasters may be either H or D like rose-comb fowls. Hence marriages between two non-tasters can only be $R \times R$ and can only have offspring R. Marriages between a taster and a non-taster may be $H \times R$ having some non-tasters among their progeny or $D \times R$ having none. Marriages between two tasters may be $H \times H$ having some non-tasters among their progeny and $H \times D$ or $D \times D$ having none. To predict what proportions of non-tasters will occur through unions between two tasters or between a taster and a non-taster, we need to know the proportions of tasters belonging respectively to the classes H and D.

This is done in the following way. Using the Principle of Random Mating, we may put the proportions in the population as:

Non-tasters .. .. a^2

Tasters (heterozygous) .. ... $2a(1 - a)$

Tasters (homozygous) .. .. $(1 - a)^2$

* Snyder (1932), "Studies in Human Inheritance. IX: The Inheritance of Taste Deficiency in Man," *Ohio Journal of Science*, XXXII.

The two kinds of marriages between a non-taster and a taster, according as the latter is heterozygous (H) or homozygous (D), will occur in the ratio:

$$2a(1 - a) : (1 - a)^2 \quad \text{or} \quad 2a : 1 - a$$

Hence, out of all marriages between a taster and a non-taster, $\dfrac{2a}{1 + a}$ belong to the first type and $\dfrac{1 - a}{1 + a}$ belong to the second. Only marriages between a heterozygous taster and a non-taster will have offspring who are non-tasters, and one half of the offspring of such marriages should belong to each class. Hence the proportion of non-tasters among the offspring of marriages between a taster and a non-taster is

$$\frac{a}{1 + a}$$

If non-tasters are 25 per cent. of the population, $a^2 = \frac{1}{4}$ and $a = \frac{1}{2}$. The expected proportion of non-tasters among offspring of marriages between a taster and a non-taster is then $\frac{1}{3}$.

If our hypothesis is correct, it has been pointed out that marriages between two tasters may be of three kinds ($H \times H$, $H \times D$, $D \times D$). When mating is random, these occur in the following proportions:

$H \times H$	$4a^2(1 - a)^2$ or	$4a^2$
$H \times D$	$4a(1 - a)^3$	$4a(1 - a)$
$D \times D$	$(1 - a)^4$	$(1 - a)^2$

The sum of the right-hand column is $(1 + a)^2$, so that of marriages between two tasters $\dfrac{4a^2}{(1 + a)^2}$ are of the type $H \times H$. One-quarter of the offspring of this type of marriage should be non-tasters, and none of the offspring of the other two types can be non-tasters. Hence the proportion of non-tasters among offspring of marriages between two tasters should be

$$\left(\frac{a}{1 + a}\right)^2$$

So if the frequency of the recessive gene is one-half as before, the expectation of recessives among the offspring of parents both belonging to the dominant class will be $\frac{1}{9}$. The very close

agreement between hypothesis and recorded data is shown by the following table of Snyder's observations which are analysed in this way. The actual frequency of non-tasters in the American community studied by Snyder was 29·8 per cent., so that $a = 0·545$.*

TABLE I

Summary of 800 families studied for inheritance of Taste Deficiency for Phenyl-thio-carbamide, showing observed and calculated proportions of tasters and non-tasters in the offspring of the various types of matings. Total for 3,643 parents and children: Tasters, 70·2 per cent.; Non-Tasters, 29·8 per cent.

Matings	Children	
	Tasters	Non-tasters
Taster × Taster, 425 ..	929 obs. 0·877 ± 0·007 calc. 0·876 ± 0·001 dev. 0·001 ± 0·007	130 0·123 ± 0·007 0·124 ± 0·001 0·001 ± 0·007
Taster × Taste deficient, 289	483 obs. 0·634 ± 0·012 calc. 0·646 ± 0·002 dev. 0·012 ± 0·012	278 0·366 ± 0·012 0·354 ± 0·002 0·012 ± 0·012
Taste deficient × Taste deficient, 86	5 obs. 0·021 calc. 0·000 dev. 0·021	218 0·979 1·000 0·021

Our knowledge of the inheritance of the isoagglutinin reaction is less recent and based on much more extensive data than the preceding analysis of taste blindness. The serum of some individuals has the property of curdling the blood of others by agglutination of the corpuscles. The reaction depends upon specific substances in the serum and in the corpuscles themselves. According to what kind of serum agglutinin and what kind of corpuscular agglutinogen a person possesses, populations may be divided into four classes. Nowadays they are called O, A, B and AB. Matings of two individuals belonging to group O only give offspring belonging to group

* For statistical constants, see Appendix I.

O. The number of exceptions to this rule are attributable to doubtful paternity. Matings of $O \times A$ or $A \times A$ can give offspring of either A or O, and matings between $O \times B$ or $B \times B$ give offspring B or O. This suggests that individuals of groups A and B are distinguished from individuals of group O by different dominant genes. Individuals belonging to group AB arise in crosses of the type $A \times B$. With a very small proportion of exceptions it is the rule that matings of the type $AB \times O$ have offspring belonging either to group A or to group B. This is only explicable on one hypothesis, which is that the groups A and B have each arisen by different substitutions for the same gene R. This means that there are six genotypes involved in the four blood groups.

$$O = RR$$
$$A = AR \text{ and } AA$$
$$B = BR \text{ and } BB$$
$$AB = AB$$

If this hypothesis is correct, matings between AB and A can yield A, B or AB, and matings between AB and B can yield A, B or AB. All recorded matings up to date are summarised in the table opposite, which is taken from Lattes.

A simple extension of·the method previously employed in considering single gene substitutions shows that random mating leads to the following proportions of genotypes, when a triple allelomorph system is involved:

RR	AR	AA	BR	BB	AB
r^2	$2ar$	a^2	$2br$	b^2	$2ab$

Here a, b, r are the gene frequencies of the three allelomorphic genes and $(a + b + r) = 1$. For any population we can easily calculate a and b as follows. If we use the symbols for the groups themselves to signify their frequencies,

$$r = \sqrt{O}$$
$$A + O = (a + r)^2 = (1 - b)^2$$

Therefore

$$b = 1 - \sqrt{O + A}$$

and

$$a = 1 - \sqrt{O + B}$$

The sum of the gene frequencies *a*, *b*, *r* calculated from these formulae should be unity. This may be used to test how far the

TABLE II

LATTES' GENERAL TABLE OF BLOOD-GROUP INHERITANCE

[*Exceptions surrounded by a circle—mostly early cases, some known to be illegitimate*]

Parental Combination	Number of Families	Number of Children in each Group			
		O	*A*	*B*	*AB*
O × *O*	1,192	2,630	(15)	(2)	—
A × *A*	1,256	476	2,364	(1)	(1)
O × *A*	2,535	2,256	3,021	(18)	(9)
B × *B*	293	126	—	532	(1)
O × *B*	997	958	11	1,230	1
A × *B*	1,104	401	791	641	580
O × *AB*	465	(38)	571	525	(34)
A × *AB*	481	(21)	525	253	307
B × *AB*	327	(13)	121	306	159
AB × *AB*	67	—	39	42	70
Total ..	8,717	6,919	7,458	3,550	1,162
		19,089			

distribution of the blood groups in different communities conform to the principle of random mating. The next table, which gives a representative sample of results obtained by different observers for

TABLE III

(FROM LATTES)

Nationality	Author	Number	O	A	B	AB	a	b	r	$(a \times b \times r)$
White Americans ..	Sanford	3,000	44·5	42·3	8·7	4·5	0·271	0·069	0·667	1·007
Danes	Sand et al. ..	1,759	42·8	42·4	11·3	3·5	0·263	0·078	0·655	0·996
Eskimaux ..	Bay-Schmidt	484	41·1	53·8	3·5	1·4	0·345	0·030	0·062	1·000
Javanese	Bais and Verhoef	1,346	39·9	25·7	29·0	5·4	0·171	0·191	0·630	0·992
Chinese	Li-Chi-Pan ..	1,500	31·3	38·1	20·7	9·9	0·279	0·167	0·559	1·005
Peru Indians ..	Arce Larreta ..	1,372	75·1	14·1	7·2	3·6	0·093	0·056	0·867	1·016
Hindus	Hirzfeld	1,000	31·3	19·0	41·2	8·5	0·149	0·291	0·560	1·000

different populations with widely divergent proportions of the four blood groups, shows that the correspondence between observation and hypothesis is very remarkable.*

§ 4

Though correlations of husband and wife enable us to determine the extent to which assortive mating occurs, the mathematical treatment of assortative mating does not assist materially toward a solution of the type of problem with which we have dealt in the preceding section unless we know the mutation rate. This is because a population does not attain equilibrium when assortative mating occurs. The effect of assortative mating, i.e. a tendency of like to mate with like, is to diminish the proportion of heterozygotes. For a single gene substitution this statement requires no mathematical proof. It is evident from the following considerations. Matings of the type $R \times R$ or $D \times D$ no not change the structure of a population because the offspring are the same as the parent. The same applies to matings $R \times H$ or $D \times H$. Since half the offspring in each case are heterozygotes and the other half like the homozygous parents, the resulting population has the same proportion as the parent population. The two types of matings which vitally affect the structure of a population are $R \times D$ and $H \times H$. The first type of mating produces offspring which are all heterozygous and therefore unlike both their parents. The other produces offspring of which one-half are either R or D and are therefore like neither parent. When mating occurs at random the production of new heterozygotes by matings $R \times D$ compensates for the loss of heterozygotes by segregation when two heterozygotes mate. If there is a tendency for like to breed with like, matings of the type $R \times D$ are proportionately diminished. Hence the production of new heterozygotes fails to balance the dissipation of heterozygotes by interbreeding of the heterozygous type. If like were attracted to unlike, the reverse would be true. Assortative mating in this sense is rare, if it occurs at all. From these considerations it follows that the proportion of heterozygous types is generally highest in a system of random mating. We may use the Principle of Random Mating as an upper limit to our calculations. Bearing this

* Bernstein (1930), "Ueber der Erblichkeit der Blutgruppen," *Zeit. f. Indukt. Abstamm. und Vererblehre*, LIV.

D

in mind, we may now deal with some applications of the Principle of Random Mating to the analysis of rare gene substitutions and with the limitations to which such treatment is subject.

The first concerns the analysis of what medical writers call "hereditary diseases." Given the appropriate environment, a condition determined by a dominant gene substitution can only occur among unaffected parents as the result of mutation. Hence it will generally be found that affected individuals have an affected parent. Consequently physical traits or diseases which are referable to dominant mutant genes will be traceable through a family pedigree without any break between successive generations. If mating occurs at random, further analysis of such conditions offers no difficulty when they are rare. Matings between two affected individuals will hardly ever occur, and, owing to the high ratio of the heterozygous to the homozygous class, practically all matings between an unaffected and an affected individual will belong to the class $R \times H$. The frequency of the matings $R \times H$ and $R \times D$ are in the ratio

$$2a : 1 - a$$

where $(1 - a)$ is the frequency of the dominant gene. Suppose that the disease has a frequency $1 : 10,000$. Matings of the type $R \times H$ are about fifty times as common as $R \times D$. Therefore the error is small if we assume that all matings between affected and unaffected parents are of this type. The expectation of offspring in a cross of the type $R \times H$ is one-half affected and one-half unaffected. If we take all the progeny of matings between an affected and a normal parent (irrespective of whether they produce any affected offspring or not), one-half the total should be affected. The error will always be very small for traits more rare than $1 : 10,000$, and the standard deviation may be determined by the formula $\sqrt{npq}$ or $\frac{1}{2}\sqrt{n}$, where n is the total number of offspring. This method yields very satisfactory results for the analysis of recorded pedigrees of such conditions as diabetes insipidus, lobster claw, major brachydactyly, aniridia, Huntingdon's chorea. An analysis of the pedigrees in the *Treasury of Human inheritance* shows that the error, though generally small, is consistently such as to favour excess of the pathotype in these cases. This will be seen from the table opposite.

Though the excess may not be significant, it is of interest to note that either of two reasons which conflict with the Principle of Random Mating in its most rigorous form would have this effect.

The first is that there may be a tendency for affected persons to marry one another. This is not conspicuously borne out by the recorded data and may be neglected. The second is that rare conditions are often highly localised. Hence the concentration of a rare gene may be *effectively* much higher than would be indicated by vital statistics. The last point may be put in another way. If few people number among their marriageable acquaintances more than a thousand, the probability of marriage with a person affected with a rare complaint or deformity is not decreased proportionately if the rarity of the condition is diminished from one in a thousand to one in a hundred thousand.

The theory of random mating leads to results of more general

TABLE IV

	Observed Number of Pathotypes	Calculated
Diabetes insipidus	79	73 ± 6
Lobster claw	114	95 ± 7
Brachydactyly	100	92 ± 7
Aniridia	170	150 ± 9

application to the propagation of rare recessive conditions. If a condition depends upon a recessive gene substitution, affected individuals may arise from three types of matings. When mating is random these occur in the following proportions:

$$R \times R \qquad a^2$$
$$R \times H \qquad 4a(1 - a)$$
$$H \times H \qquad 4(1 - a)^2$$

It is clear that if the recessive type is rare (i.e. a^2 is a very small fraction), recessive offspring will rarely have a recessive parent. They will usually have two apparently "normal" parents ($H \times H$) who are carriers. If a recessive marries an individual who is not of the recessive type, the alternatives occur in the proportions:

$$R \times H \qquad 2a$$
$$R \times D \qquad 1 - a$$

Since recessives are only produced by matings of the first type, it follows that recessive parents rarely have recessive offspring. So if

rare, recessives are rarely the parents or the offspring of recessives. Hence they will not be readily traced in successive generations of a pedigree. It is necessary to adopt special methods to recognise them as such.

If there is a tendency for like to breed with like, large departures from the results predicted according to the Principle of Random Mating are much less likely when we are dealing with rare recessive conditions than when rare dominants are involved. This is easily shown to be true in the limit. It is not necessary to do so in this context, because the reason can be grasped without recourse to symbols. The effect of assortative mating is to increase matings of the types $D \times D$, $H \times H$, $D \times H$ and $R \times R$ and to diminish $R \times D$ or $R \times H$. If D is small compared with H, the concentration of matings $D \times D$ and $D \times H$ will be small compared with $H \times H$. In other words, the tendency for H to disappear will be great. If R is small compared with H, the concentration of matings of the type $D \times H$ or $D \times D$ will be great compared with $H \times H$. In other words, the tendency for H to disappear will be relatively less.

Departures from the Principle of Random Mating applied to rare recessive conditions can be tested by means of a relation already given. This will be illustrated by the case of albinism. Albinism has a frequency of about $1 : 20,000$ in European countries. As will be seen in Chapter III, it depends on a recessive gene substitution. If mating occurs at random, marriages between (a) an albino and a carrier, (b) an albino and a normal person who does not carry the gene, will occur in the ratio

$$2a : 1 - a$$

Since matings of the second kind do not produce albino offspring, the proportion of marriages between albinos and persons who are not albinos but may have albino offspring would be

$$\frac{2a}{1 + a} = \frac{1}{70} \text{ (approximately)}$$

If positive assortative mating occurs, the proportion of heterozygous individuals will be less than if mating occurs at random. The effect of this would be to diminish the proportion of marriages of the type $R \times H$ as compared with $R \times D$ still further. To test the frequency with which albinos may have albino offspring when married to individuals who are not albinos, it is necessary to make

an adjustment on account of the small size of human families. In a mating of the type $R \times H$, one-half the offspring will be normal. If a family contains s individuals, the probability that all of them will be normal is $(\frac{1}{2})^s$. Thus only a proportion $1 - (\frac{1}{2})^s$ of all families which might have some albino offspring actually have at least one, and are distinguishable from families produced by matings of the type $R \times D$. If there are n_1 families of one member, n_2 of two . . ., n_s of s members with one albino and one normal parent, the proportion of such families with any albino offspring will be:

$$\frac{2a}{1 + a} \cdot \frac{\Sigma(1 - (\frac{1}{2})^s) \cdot n_s}{\Sigma n_s}$$

This will be appreciably less than $\dfrac{1}{70}$ if the sample contains a number of very small families. The next table shows an analysis of matings of this type in Pearson's monograph which contains several hundred pedigrees in which albinism occurs.

TABLE V

s	Families with no Affected Offspring	Families with at least one Affected Offspring	Total (n)	$n_s[1 - (\frac{1}{2})^s]$
1	4	1	5	2·500
2	2	0	2	1·500
3	1	0	1	0·875
4	4	2	6	5·625
6	4	2	6	5·906
7	1	1	2	1·948
8	3	2	5	4·980
9	4	0	4	3·992
10	2	0	2	1·998
12	1	0	1	1·000
Total ..	26	8	34	30·360

According to the figures in this table, the expected proportion of families with at least one albino offspring is

$$\frac{1}{70} \times \frac{30·4}{34} = \frac{1}{79}$$

Actually the observed proportion is $\dfrac{8}{34}$, or nearly twenty times as great as it should be. This cannot be due to the tendency of albinos to marry their own kind, because assortative mating in that sense would have the opposite effect. One obvious explanation is that in using the figure $\dfrac{1}{70}$ based upon the net frequency of albinos, we assume that the gene is distributed uniformly in the population. In fact this is a most unlikely assumption when we are dealing with a rare gene. Rare genes, especially if selection is unfavourable to their survival, will tend to be isolated in local pockets of the population. Hence the *effective* frequency of albinism may be very much higher in regions where albinos are found to be living.

This emphasises a conclusion already mentioned. Correlations of married couples selected over a wide area may give the impression that a high degree of assortative mating occurs in the population. For instance, if we take pairs at random from different villages in Great Britain, it is not unlikely that we should find a noticeable correspondence in eye colour, because small inbred communities contain a high proportion of individuals of any given type. This is fully consistent with the absence of any tendency to assortative mating when each individual has equal access to individuals of any given type. Assortative mating in the latter sense decreases the likelihood that a recessive will have recessive offspring if married to an individual who is not a recessive. Geographical isolation increases the probability that a recessive will have recessive offspring if married to an individual who is not a recessive. If the use of aeroplanes ever becomes universal, the Principle of Random Mating will be more suitable to the study of rare gene substitutions.

The limitations imposed by geographical propinquity are well seen in connexion with sex-linked conditions such as haemophilia. This is a very rare disease. In a large community like Great Britain, it has an incidence lower than 1 in 100,000. Matings of haemophiliacs with healthy women who are (*a*) carriers, (*b*) not carriers occur with the frequency

$$RY \times RD \qquad \tfrac{1}{2}a^2(1 - a)$$
$$RY \times DD \qquad \tfrac{1}{4}a(1 - a)^2$$

Out of all marriages between haemophiliacs and healthy women, the proportion of marriages between a haemophiliac and a carrier is

$$\frac{2a}{1 + a}$$

In all pedigrees published, about $\frac{1}{25}$ marriages between a haemophiliac and a healthy woman produce haemophiliac offspring. This is a very high proportion compared with what it would be if haemophilia were distributed evenly throughout the population. In fact, many haemophiliac pedigrees are taken from small isolated communities where the frequency of the recessive gene may be high.

SUMMARY

The Principle of Random Mating can be used to make quantitatively verifiable predictions concerning single gene substitutions and multiple allelomorph systems, when the several genotypes are (*a*) not distinguishable in such a way as to affect the choice of partners; and (*b*) present in the population in proportions which do not differ very greatly. When these conditions are not realised, the Principle of Random Mating draws attention to several important conclusions which are qualitatively true. One is that rare recessives are rarely the offspring of recessive parents. More important still is the fact that consanguineous parentage is common among rare recessives. This will be deduced in the ensuing chapter, and examples drawn from the study of familial diseases will then be given.

CONSANGUINEOUS PARENTAGE AND THE THEORY OF INBREEDING

§ I

One important conclusion which emerges from a consideration of the results of random mating is that rare recessives are rarely the offspring or parents of recessives. This might seem to present an insuperable obstacle to their recognition. The detection of single recessive gene substitutions no longer offers any special difficulty. One method of recognising them depends upon the theory of inbreeding. The modern theory of inbreeding based upon the particulate nature of hereditary transmission leads to an important conclusion which can be deduced in algebraic form. Stated in general terms it is this, *When a recessive trait is rare, a large proportion of individuals who exhibit it are the offspring of unions· between near relatives.* This means that a systematic study of the issue of consanguineous unions carried out on a large scale may very considerably extend the list of known recessive mutations in human beings.

The criterion of consanguinity was first stated by Garrod in a paper on alcaptonuria. It was published in the *Lancet* in the same year as the application of Mendel's laws to animals by Bateson. Alcaptonuria is a familial disease in which the incidence of consanguinity is particularly high. That is why alcapton, the highly oxidisable urine constituent of alcaptonuric cases, has earned the name "homogenistic" acid. The percentage of cousin marriages in European populations generally lies between 0·5 and 1 per cent. of all marriages. In rural areas, among the landed aristocracy, and · in Jewish communities it may be higher. An upper limit of 3 per cent. would probably cover the more extreme cases. The latest review of alcaptonuria shows that there are 151 cases of the disease recorded in medical literature. Of these, the facts about the family histories of 83 are recorded. The results up to date are summarised in the table opposite.

In the table, 27 out of 63 alcaptonurics concerning whose parentage precise information is available are offspring of· consan-

guineous parents. In all except one of these the parents were first-cousins. This would make the percentage of consanguineous parentage 43 per cent. and of first-cousin marriages among parents of alcaptonurics 42 per cent. One isolated case is also known to be the offspring of related parents. Thus 28 out of the 151 recorded cases are known to be the issue of inbred unions. This means that the percentage of consanguineous parents among parents of alcaptonuric offspring is at the most conservative estimate at least 18 per cent., or twenty times as high as the percentage of cousin marriages among all patients in London hospitals.

Garrod grasped the significance of this excess by very simple

TABLE VI

ALCAPTONURIA

Relationship of Parents	Number of Fraternities	Alcaptonurics		Total
		Female	Male	
First-cousins	12	7	19	26
Consanguineous (degree unknown)	1	0	1	1
Not ascertained ..	12	6	14	20
Unrelated	20	15	21	36
Total	45	28	55	83

reasoning, which was mentioned in the first chapter. Nearly all known cases of alcaptonurics are offspring of parents who do not share the disorder. To be precise, only 10 out of 138 have an affected parent, if we exclude an anomalous pedigree of Pieter. If we assume a genetic basis to account for two outstanding features of alcaptonuria, namely, its familial character and the peculiarity under discussion, parents of alcaptonurics may be regarded as "carriers," or, in more precise terms, as heterozygotes. If an individual is a heterozygote, his cousin is more likely to be a heterozygote than would another individual taken at random from the community. Garrod argued correctly that cousin unions increase the likelihood that two heterozygous types will marry, and hence a high proportion of the parents of a rare recessive will be first-cousins. His paper is

a landmark in the history of human genetics. It attracted comparatively little notice at the time of publication. Before its importance could be fully appreciated, it was necessary to formulate the principle involved in more precise terms.

The approximate treatment of the concentration of recessives in first-cousin marriages as given by Lenz involves the least difficulty to the student who is not a mathematician. It does not apply to sex-linked gene substitutions. It assumes that random mating occurs, and that a condition is so rare that only marriages between carriers need to be taken into account. If mating is at random, this
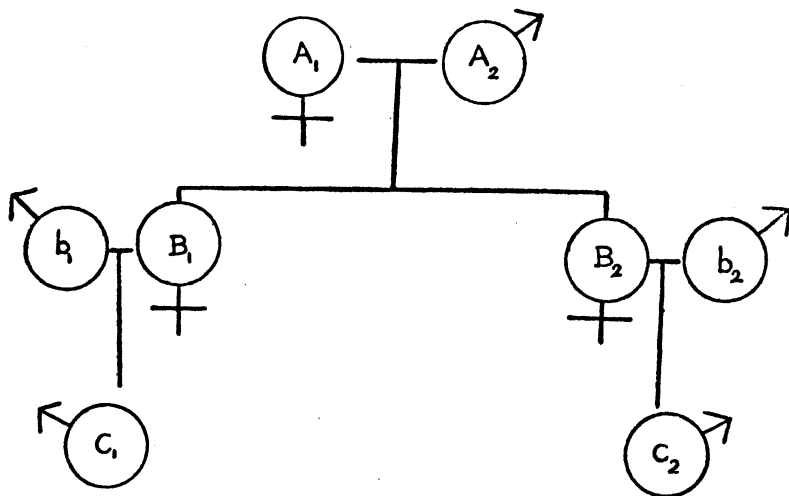
FIG. 1

does not introduce a very serious error providing the incidence is less than about 1 in 2,000. The reasoning will be assisted by the help of the accompanying diagram (Fig. 1), which shows two boy cousins C_1 and C_2, whose mothers B_1 and B_2 are daughters of the common grandparents A_1 and A_2. If C_1 is a heterozygote, the chance that he gets the recessive gene from his mother (B_1) rather than his father (b_1), and, that is, from the common grandparents of C_1 and C_2, is $\frac{1}{2}$. If B_1 has the gene, either A_1 or A_2 must have it, so that the chance that B_2 has it will be $\frac{1}{2}$. The chance that B_2 has it is thus $\frac{1}{2}$ the chance that B_1 has it, or $\frac{1}{4}$ the chance that C_1 has it. The chance that C_2 has it is $\frac{1}{2}$ the chance B_2 has it, or $\frac{1}{8}$ that C_1 has it.

The frequency of cousin marriages, in other words, the proba-

bility that a man will marry his cousin, will be denoted by x. If a is the frequency of the recessive gene, the probability that two heterozygotes will marry is

$$2a(1 - a) \times 2a(1 - a) = 4a^2(1 - a)^2$$

The probability that two heterozygotes who are *not* cousins will marry is therefore

$$4a^2(1 - a)^2(1 - x)$$

Since a quarter of their offspring will be recessives, we may write for the frequency of recessives with two parents who are heterozygous and not cousins,

$$a^2(1 - a)^2(1 - x)$$

The probability that an individual is heterozygous is $2a(1 - a)$. The probability that an individual marries his cousin is x. The probability that an individual will both be heterozygous and marry his cousin is therefore $2a(1 - a) . x$. If an individual is heterozygous, the probability that his cousin will also be heterozygous is $\frac{1}{8}$. Therefore the probability that an individual will both marry his cousin who is a heterozygote and himself be a heterozygote is

$$[2a(1 - a) . x] \times \tfrac{1}{8} = \tfrac{1}{4}ax(1 - a)$$

Since a quarter of the issue of such a union will be recessives, the frequency of recessives who are the offspring of marriages between heterozygous individuals who are cousins is

$$\frac{ax}{16}(1 - a)$$

The proportion of recessives arising from first-cousin marriages of all kinds is therefore

$$\frac{\dfrac{ax}{16}(1 - a) \cdot}{a^2(1 - a)^2(1 - x) + \dfrac{ax}{16}(1 - a)} = \frac{x}{x + 16a(1 - a)(1 - x)}$$

Since x is always small compared with unity, and a is necessarily small, when the bulk of unions belong to the class $H \times H$ this tends to become

$$\frac{x}{x + 16a}$$

If we apply this formula to the case of alcaptonuria, which has an incidence of about $\dfrac{1}{1,000,000}$, a is the square root of the frequency of the recessive type (a^2), or $\dfrac{1}{1,000}$. Taking 8 in 1,000 as the normal frequency of cousin marriages, we should expect the frequency of cousin marriages among alcaptonurics to be

$$\frac{\dfrac{8}{1,000}}{\dfrac{8}{1,000}+\dfrac{16}{1,000}} = 33 \text{ per cent.}$$

This compares very well with the observed estimate, which lies between an upper limit of 42 per cent, and a lower limit of 18 per cent., with a mean value of 30 per cent.*

Among other conditions which are associated with a percentage of first-cousin parentages of 15 or over—that is, about twenty times the usual frequency of first-cousin marriages—the following are conspicuous examples:

> Albinism.
> Retinitis pigmentosa (Usher's *familial* type).
> Xeroderma.
> Total congenital colour blindness.
> Amaurotic family idiocy.
> Maculo-cortical degeneration (juvenile form of Tay Sach's disease).
> Friedreich's ataxia.
> Congenital ichthyosis.

In addition to these, a high proportion of consanguineous parentage occurs notably in deaf-mutism.

If we restrict ourselves to the case with which this formula of Lenz deals, namely, recessives with parents who are not recessives, it will be seen that a recessive condition with a frequency of 1 in 5,000 would yield a percentage of first-cousin marriages roughly equal to 1·5 per cent. This would be too low to be regarded with confidence as a significant excess. It is an arresting and highly significant fact that consanguineous parentage is often greatly in excess of the

* Hogben, Worrall and Zieve (1932), "The Genetic Basis of Alcaptonuria," *Proc. Roy. Soc. Edin.*, Vol. LII, Part III, No. 13.

figure calculated from the Principle of Random Mating. For instance, about 17 per cent. of albinos have parents who are first-cousins. The frequency of albinism in European countries is of the order 1 in 20,000. From the formula given above, about 6·5 per cent. albinos should therefore have parents who are first-cousins. The observed proportion is nearly three times the expected figure. This is the more surprising, because we have already seen that the effective frequency of a recessive trait is much higher than the actual frequency given in the vital statistics of a population. Therefore the formula is more likely to exaggerate the contribution of consanguinity than to give an estimate which is too low.

A simple explanation put forward by Roberts appears to meet the case. Albinism may not be a simple genetic entity. There are

· TABLE VIA

AMAUROTIC FAMILY IDIOCY

	Number of Cases	Parents First-Cousins	Parents Consanguineous
		Per cent	Per cent
Jews	97	11·9–16·0	23·0–30·9
Gentiles	19	33·3–40	55·6–66·7

various black-bodied mutations in the fruit fly. "Sable" is referred to a gene substitution on the first chromosome, "black" to a gene substitution on the second and "ebony" to a gene substitution on the third. So possibly several independent recessive mutations with different loci may give rise to white hair and pink eyes in man. In that case, the frequency of any of them must be necessarily low. Supposing albinism is the manifestation of either of two genes with the same frequencies, the percentage of cousin marriages among parents of albinos calculated from the Lenz formula would be roughly 18 per cent., or three times as high as it would be on the assumption that it is a single genetic type.

It is clear that any such formula as that which has been given by Lenz is subject to the same limitation as the Principle of Random Mating from which it is deduced. Its main value is that it shows where the application of arguments derived from common sense necessarily breaks down, and, in a general way, it is borne out by the fact that very rare recessive conditions tend to show a higher

excess of consanguineous parentage than do ones which are more common. This is well illustrated by the infantile form of Tay Sach's disease. This was once believed to be an exclusively Jewish complaint. However, there are a few authentic Gentile cases on record. Dr. Slome's* recent analysis of the data shows that the incidence of consanguinity among parents of the Gentile cases is much higher than for the Jewish ones. This might be expected from their greater rarity (Table VIA).

§ 2

An exact analysis of the effect of consanguineous parentage is of interest first and foremost because it provides a very powerful instrument for the detection of recessive mutations. Its clinical interest does not end here. The physician may be asked whether it is wise for first-cousins to marry. This raises other issues of medico-legal importance. The law of this country forbids marriages between persons of a certain degree of relationship, and severely punishes sexual connexion between persons more intimately related. Thus it is illegal for a man to marry his aunt, but it is not a criminal offence for aunt and nephew to produce an illegitimate child. It is a criminal offence for a brother and sister to cohabit, whether they produce offspring or not. The constitution of Great Britain endows bishops with special prerogatives to legislate in such issues. It does not extend the same prerogatives to biologists.

The effects of consanguineous parentage may be examined from two different points of view. One is to determine what happens when all matings are carried out in a particular way; for example, breeding exclusively by brother-sister or parent-child matings. The other is to examine how different genotypes are distributed among the offspring of relatives and unrelated parents when mating occurs at random. The first problem was the basis of an important memoir by Jennings in 1916. The theory of inbreeding developed by Jennings has one most important application in the experimental branches of medical science. It leads to the conclusion that any system of close inbreeding diminishes the proportion of heterozygotes. It thus provides a criterion for the genetic homogeneity of a stock and in consequence a refinement in the technique of properly controlled observations. In human affairs it has no rele-

* Slome (1933), "The Genetic Mechanism of Amaurotic Family Idiocy," *Journ. Genetics* (in press).

vance to any contemporary problems. The main issue, as it has any bearing on medico-legal issues, is how different genotypes are concentrated among offspring of consanguineous unions, which are few in comparison with marriages of unrelated individuals.

An exact answer to this question can be given if mating occurs at random. Such an answer will be true to a high degree of approximation if the trait involved is relatively common, uniformly distributed in a community and is not one which influences the choice of a partner. Even when these conditions are not realised, the results of a purely theoretical treatment show whether one degree of consanguineous union provides a higher or lower proportion of homozygous types than another. The treatment of autosomal genes and genes which are sex-linked must be dealt with separately. If we consider autosomal gene substitutions, we find that the proportion of homozygotes, whether of the recessive or dominant type, is high and the proportion of heterozygotes low among the offspring of consanguineous unions as compared with the frequency of the homozygous and heterozygous classes in the population as a whole. With a sex-linked gene substitution this is only true of female offspring. There is no male heterozygous class, and the proportions of the two homozygous classes among offspring of consanguineous unions do not differ from that which would be found in the rest of the population.

The concentration of recessives among the offspring of unions between a parent and child for a single autosomal gene substitution can be set forth simply in tabular form. The frequencies of the six

TABLE VIB

Frequency	Matings of P_1	Offspring F. 1	Matings of $F_1 \times P_1$	Recessive Offspring
a	$R \times R$	R	RR	1
$4a^3b$	$R \times H$	$(\frac{1}{2}R \times \frac{1}{2}H)$	$\frac{1}{4}RR + \frac{1}{2}RH + \frac{1}{4}HH$	$(\frac{1}{4} + \frac{1}{4} + \frac{1}{16})$
$2a^2b^2$	$R \times D$	H	$\frac{1}{2}RH + \frac{1}{2}RD$	$\frac{1}{4}$
$4a^2b^2$	$H \times H$	$(\frac{1}{4}R + \frac{1}{2}RH + \frac{1}{4}D)$	$\frac{1}{4}RH + \frac{1}{2}HH + \frac{1}{4}DH$	$\frac{1}{8} + \frac{1}{8}$
$4ab^3$	$H \times D$	$(\frac{1}{2}H + \frac{1}{2}D)$	$\frac{1}{4}HH + \frac{1}{2}DH + \frac{1}{4}DD$	$\frac{1}{16}$
b^4	$D \times D$	D	DD	0

Total Recessive Offspring:

$$a^4 + \tfrac{9}{16} \times 4a^3b + \tfrac{1}{4} \times 2a^2b^2 + \tfrac{1}{4} \times 4a^2b^2 + \tfrac{1}{16} \times 4ab^3$$

Putting $b = 1 - a$, this becomes $\frac{1}{4}(a + 3a^2)$

NATURE AND NURTURE

possible classes of matings when the frequency of the recessive homozygote is a^2 have been given already, and are shown in the left-hand column.

In a similar way we may deduce for other types of relationship the following values for the frequency of recessives among the offspring of various kinds of consanguineous unions involving a single autosomal gene substitution when mating occurs at random and the frequency of recessives in the general population is a^2.

TABLE VII*

Unrelated	a^2
Parent-child .. .. $\left.\vphantom{\begin{matrix}a\\a\end{matrix}}\right\}$	$\frac{1}{4}(a + 3a^2)$
Brother-sister	
Uncle or aunt with niece or nephew .. .. $\left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\}$	
Half-brother with half-sister	$\frac{1}{8}(a + 7a^2)$
Grandparent with grandchild	
First-cousins	$\frac{1}{16}(a + 15a^2)$
Second-cousins	$\frac{1}{64}(a + 63a^2)$

The frequencies of the .homozygous dominant are obtained by substituting $(1 - a)$ for a. When transmission is sex-linked, the concentration of recessive females in consanguineous unions is given in Table VIII.

TABLE VIII

Unrelated	a^2
Father-daughter	$\frac{1}{2}(a + a^2)$
Mother-son	$\frac{1}{2}(a + a^2)$
Brother-sister	$\frac{1}{4}(a + 3a^2)$
Nephew with paternal aunt	a^2
Nephew with maternal aunt	$\frac{1}{8}(3a + 5a^2)$
Niece with paternal uncle	$\frac{1}{4}(a + 3a^2)$
Niece with maternal uncle	$\frac{1}{8}(a + 7a^2)$
Paternal cousins	a^2
Maternal cousins	$\frac{1}{16}(3a + 13a^2)$
Cousins ($♀$ with mother's brother's son)	a^2
Cousins ($♂$ with mother's brother's daughter) ..	$\frac{1}{8}(a + 7a^2)$

* 1. Haldane and Moshinsky (1933), "Partial Inbreeding in Man," *Journ. Genetics* (in press).

2. Hogben (1933), "A Matrix Notation for Mendelian Populations," *Proc. Roy. Soc. Edin.*, Vol. LIII.

Before discussing in what sense these conclusions point to the advisability or otherwise of a consanguineous relationship, it is interesting to note that the degree of social disfavour which pertains to different classes of inbred unions corresponds on the whole fairly closely with the extent to which they yield offspring in proportions more or less unlike those in which the several genotypes occur in the population at large. For instance, the most profound disturbance of the general genotype ratios is found in the offspring of incestuous unions, i.e. parent-and-child or brother-and-sister. Next come uncle or aunt with niece or nephew. The effect of a marriage between first-cousins is less drastic than a union between uncle or aunt and niece or nephew. If sex-linked genes are involved, such conclusions require modification. Thus, in certain circumstances, a union with a paternal uncle may involve consequences of the same order as marrying a sister, whereas a union with a paternal aunt may involve no deviation from mating at random or marrying a deceased wife's sister.

From a clinical point of view these deductions give rise to two questions. The first is whether it is advisable to encourage marriage between near relatives in the interests of individual parents. The second is, what would be the result to the community as a whole if consanguineous marriages were encouraged more than we do at present? As regards the first, two classes of considerations have to be accommodated. By increasing the chance of having children with a rare recessive condition or a condition determined by a combination of rare recessive genes, we may produce a peculiarly gifted or a peculiarly defective offspring. In the present state of knowledge it would be unwise to state dogmatically that the latter is the more likely result. At the same time, the risk is great for several reasons. It may be argued that, although the chance of having a child who is a juvenile amaurotic is much larger in a cousin marriage, it is still a small chance. Against this it is well to remember that the chance of having one or other of a large class of recessive disorders is not necessarily a small one. The vital question is whether the class is large or small. At present the list of recessive mutations in man is a small one. This may be because we have only recently perfected the means of detecting them. Further evidence for this view will be discussed in the ensuing chapter. In the meantime two general conclusions are relevant. The first is that recessive mutations are far more common than the dominant type in animals. The

second is that recessive mutations are weeded out very slowly by selection when they are rare. Here consanguineous matings will tend to encourage the appearance of lethal and semi-lethal types which are normally protected from selection by the small chance one carrier has of meeting another. For these reasons the risk is great. It is doubtful whether the average person would take it if he were conversant with its magnitude.

An additional complication arises when we discuss the encouragement of consanguineous unions from the standpoint of the community as a whole. The enthusiastic eugenist might be disposed to argue as follows. By encouraging the appearance of pure types, inbreeding favours the rapid elimination of lethal and semi-lethal conditions. It also favours the emergence of rare combinations of genes whose combined effect may be socially valuable. Curiously enough, the last statement is not entirely true, for a reason which has been recently pointed out by Haldane and Waddington in a mathematical analysis of the effect of inbreeding upon linked genes. The algebraical treatment of the problem, though not difficult, is very laborious. So it must suffice to summarise their argument.

Supposing there arise in a population a series of genes l, m, n, etc., all located on the same pair of chromosomes. If mating occurs at random, the result of repeated crossing over is that l and m will occur with equal frequencies in the same or alternative members of the pair to which they belong, when they are both present in the zygote. Hence every possible combination of linked genes is eventually reached. If assortative mating or inbreeding occurs, the course of events is different. Crossing over only affects the redistribution of different genes in the heterozygous condition. Since inbreeding progressively diminishes the proportion of heterozygotes in a population, it limits the tendency to bring about new combinations of linked genes.*

<div style="text-align:center">´ §3</div>

In what has been said so far attention has been focussed mainly upon the results of inbreeding as it affects the appearance of clearcut disorders such as alcaptonuria or amaurotic family idiocy. There is a common belief that consanguineous unions encourage feeblemindedness and insanity. It is difficult to say whether this is so,

* Haldane and Waddington (1931), "Inbreeding and Linkage," *Journ. Genetics*, Vol. XVI.   ,

because the existing data suffer from lack of any uniform system of classification. In all classes of mental defectives or insane patients the percentage of consanguineous parentage is somewhat higher than for the general population. It is not so high as to make it certainly higher than local populations from which such data are taken. Until a more refined clinical classification is agreed upon, little progress along these lines is likely to be made. In any case, simple primary amentia, if it is really as simple a category as the people who are placed in it, is not very rare. That is to say, it is not so rare that we should expect to obtain a very significant excess of consanguineous parentage among children who are classified in that way. The belief that the milder mental disorders are likely to be recognised as "unit characters" is a survival of the instinct psychology of Spencer's period, and it has little support from modern researches into the physiology of the central nervous system.

There is more than one way in which the theory of inbreeding might be used to throw light on the contribution of genetic differences to variability of social behaviour. The fact that consanguinity increases the chance that rare recessives will make their appearance is only a particular case of a more general law. This is that the variability of a group of individuals produced by related parents is greater than the variability of a group of individuals produced by unrelated parents. The usual measure of variability is the mean square deviation from the mean value of a metrical character. Fisher calls this the *variance* of a population.

First, consider a character whose measurements are determined by a single gene substitution, making the unit of measurement the difference between the two homozygous genotypes R and D. If the heterozygous type differs from R by a fraction A of the unit step between R and D, the representative measurements of R, H, D are X, $X + A$, and $X + 1$. When the origin of measurement is at X, the measurements become 0, A, 1 respectively. Using R, H and D for the frequencies of the genotypes in the population, the mean value is

$$(A)H + D$$

The mean square deviation measured from the origin* is

$$(A)^2 H + D$$

* When the mean square deviation is computed from the origin of measurement, the mean square deviation (σ^2) from the mean is given by

$$\sigma^2 = \sigma_0{}^2 - M^2$$

where M is the mean.

Measured from the mean it is

$$A^2H + D - [(A)H + D]^2 = V$$

The ratio of the variance of offspring of cousins (V_C) and offspring of unrelated individuals (V_R) can be found at once by substituting the appropriate values of R, H, D, viz.:

	Offspring of Unrelated Parents	Offspring of Cousins
R	a^2	$\dfrac{a}{16}(1 + 15a)$
H	$2ab$	$\dfrac{30}{16}ab$
D	b^2	$\dfrac{b}{16}(1 + 15b)$

When the heterozygote is exactly intermediate between the two homozygous types, $A = \frac{1}{2}$; and

$$\frac{V_C}{V_R} = \frac{17}{16}$$

If the heterozygote has the same measurement as one of the homozygous (D) classes, $A = 1$; and

$$\frac{V_C}{V_R} = \frac{(16 + 15a)(16 - 15b)}{256a(1 + a)}$$

The value of the expression on the right has a minimum of $\dfrac{779}{768}$ when $a = \frac{1}{2}$, and becomes larger as a diminishes. When the recessive genotype has a frequency of $1 : 100$ (i.e. when $a = 0 \cdot 1$) it is roughly $\frac{3}{2}$. When $a = 0 \cdot 01$ it lies between 6 and 7.

For the two conditions specified, i.e. when the heterozygous type is identical with the dominant class, or exactly intermediate between R and D, the same formulae hold good for a range of measurements determined by n pairs of genes having the same frequency $(a_1 = a_2 = a_1,$ etc.). This is easily seen when two pairs of genes determine the measured character, and the effects of each gene

are additive and equivalent. The genotypes may be represented thus:

Genotype	Deviation	Frequency of Phenotupes when $R_1 = R_2$, etc.
$a_1a_1a_2a_2$	0	R_1R_1
$a_1A_1a_2a_2 \quad a_1a_1A_2a_2$	$\frac{1}{4}$	$2R^1H_1$
$a_1a_1A^2A^2 \quad A_1A_1a_2a_2 \quad A_1a_1A_2A_2$	$\frac{1}{2}$	$H_1H_2, \ 2R_1D_1$
$A_1A_1a_2A_2 \quad A_1a_1A_2A_2$	$\frac{3}{4}$	$2H_1D_1$
$A_1A_1A_2A_2$	1	D_1D_1

It is at once seen that the mean is:

$$\tfrac{1}{2}R_1H_1 + \tfrac{1}{2}H_1H_1 + R_1D_1 + \tfrac{3}{2}H_1D_1 + D_1D_1$$
$$= \tfrac{1}{2}H_1(R_1 + D_1 + H_1) + D_1(R_1 + D_1 + H_1)$$
$$= \tfrac{1}{2}H_1 + D_1$$

Similarly the variance is found to be

$$\tfrac{1}{8}H_1(1 - H_1) + \tfrac{1}{2}R_1D_1$$

On substituting the appropriate values for R, H and D, we obtain, as before,

$$\frac{V_C}{V_R} = \frac{17}{16}$$

If we call the quantity $V_C : V_R$ the Variance Ratio, the following conclusions hold good for metrical characters: (i) When the observed measurements are determined by genes whose effect is independent, perfectly additive and equivalent, the variance ratio is not affected by the number of pairs involved or their frequencies. (ii) When the effect of the genes is not strictly additive, the variance ratio depends both upon the degree of dominance and the frequency of the genes involved.

A systematic study of the variance with respect to the results of intelligence tests among children of parents who are first-cousins and children of unrelated parents might considerably add to our knowledge of the part which hereditary differences play in determining differences of "intelligence." In a mixed population the distribution of intelligence quotients approximately follows a normal

curve. This might result from a combination of a small number of genetic differences and large discrepancies between environmental agencies acting as limiting factors to their manifestation. It might also mean that a very large number of gene differences, affected very little by differences of the general environment in which they are situated, determine the distribution of intelligence in a population. Eugenists often take it for granted that the latter is true. At present there is insufficient evidence for such a belief. Likewise there is no *prima facie* reason for the view that what are called diseases of the mind are less numerous than diseases of the body, that heredity plays a more important rôle in the manifestation of one than of the other, or that we have less reason to look to control of the environment as a means of remedying one than the other. There is little doubt that individuals differ genetically with respect to their liability to be born idiots or to die of heart disease. There is perhaps as much reason to seek a remedy for one as for the other.

The normal distribution of intelligence-test scores in a population makes it impossible to attribute the large class of simple primary amentia to any single or simple form of hereditary transmission. This is borne out by other considerations. Pedigrees in which "simple" primary amentia regularly reappears as a *dominant* condition in successive generations are the exception rather than the rule. It is also true that a high familial incidence is not common. If a condition is determined by a single recessive gene substitution, it will generally manifest itself among several sibs, because the probability that an individual will be recessive if one of his sibs is recessive is a quarter. So if simple primary amentia is a single clinical entity, its appearance is either determined by several genes or by a single gene substitution which only manifests itself in a special kind of environment not necessarily common to all members of a fraternity. It is highly unlikely that the arbitrary delimitation of a subcultural group with an IQ below 75 and no special clinical peculiarities defines one and the same biological phenomenon. It is not impossible that such a group does contain classes of individuals whose deviation from the standard is mainly due to a single recessive gene substitution. In a recent investigation, Dr. Penrose* has made use of another application of the modern theory of consanguineous parentage to test this possibility.

If such a class of individuals exists, we are entitled to conclude

* Private communication to the writer.

that they make an unduly large contribution to that section of the mentally defective population derived from consanguineous unions. This would imply that the presence of several affected members in the same fraternity would occur more commonly in fraternities with consanguineous parents than in fraternities with unrelated parents. Dr. Penrose found that this is conspicuously true. Indeed, the familial incidence of mental defect in a group of fraternities with at least one defective sib having parents who were first-cousins was in close agreement with the figure predicted by the method which will form the subject-matter of the next chapter. Analysis of this kind shows how it is possible to take the hotly contested problem of mental defect out of the realm of partisanship. As with diseases of the body, so with mental diseases, different mutations and different kinds of environment may in different situations be chiefly responsible for what is described as one and the same clinical entity. Mental disease is not a single genetic problem. It is not a genetic problem alone. The study of how gene differences in one environment contribute to mental disorders does not preclude the study of other kinds of environment in which such differences would not be detected. The fact that such differences might be slowly eliminated by genetic control is quite consistent with the possibility that mental defect will eventually yield to curative treatment.

SUMMARY

When an autosomal recessive gene is rare, the probability that it will be present in duplicate among some of the offspring of parents who are near relatives is much greater than that it will be duplicated among the offspring of unrelated parents. This explains the high incidence of consanguineous parentage among persons suffering from rare familial diseases.

As a means of detecting single recessive gene substitutions, this method is not limited to the study of gene differences which are · manifest throughout the whole range of environment to which members of the same fraternity are normally exposed.

For metrical characters this conclusion signifies that the variance among offspring of related parents is greater than the variance among unrelated parents drawn from the same population.

THE GENETIC ANALYSIS OF FAMILIAL DISEASES

§ 1

An examination of the Principle of Random Mating led us to the conclusion that rare recessive traits rarely appear in the parents or offspring of individuals who exhibit them. On the other hand, they tend to reappear in the same fraternities. This is so for the following reason. If an individual is a recessive and has two parents who do not exhibit the recessive trait, both parents must be heterozygous. On the average, one-quarter of their offspring will therefore be recessive. Thus diseases determined by rare recessive genes belong to the class which are called by clinicians "familial," in contradistinction to "hereditary," diseases. Not all familial diseases are determined in this way. If a familial disease is more rare than about 1 : 5,000, it can usually be recognised as the manifestation of a recessive gene substitution by excess of consanguinity among parents of individuals suffering from it. If it is less rare than 1 : 5,000, this criterion is of little use. This chapter will deal with methods of analysis which are not subject to this limitation.

The quantitative verification of genetic hypotheses applied to human data only presupposes a knowledge of the phenotypic characteristics of a large group of parents and their offspring. In the past, students of human hereditary have devoted their efforts very largely to the collection of long pedigrees illustrating the recurrence of striking individual differences. The bulk of material of this kind is derived from medical literature. The collection of pedigrees has had an important use. As Gates and others have emphasised, similar conditions may be determined by different genes. There are at least three kinds of white plumage in the domestic fowl. One has arisen as a dominant sex-linked mutation, one as a dominant mutation which is not sex-linked and a third appears to be recessive. A similar pink eye colour in *Drosophila* is produced by genes located at several different places in the X chromosome and in addition by genes in both the second and third chromosomes. The clinical condition of congenital night blindness is an analogous case. A cursory examination of the recorded pedigrees shows that one

hypothesis could not possibly fit the facts. There is one type which is dominant and not sex-linked, one which is recessive and sex-linked, and possibly a third recessive type which is not sex-linked.

However, it must not be thought that a cursory examination of pedigrees suffices to decide whether they are consistent with a single hypothesis. For instance, a group of pedigrees may contain some families in which a trait appears to be familial and others in which it appears to be hereditary. Only statistical analysis can then decide between two alternative interpretations. The first is that the same trait may be due either to a single dominant or to a single recessive gene substitution. The second is that the trait is determined by the interaction of two independent dominant genes. It requires refined methods of quantitative analysis to distinglish between these two possibilities. In any case, long pedigrees are not necessary to determine whether data derived from two generations are consistent with a single hypothesis when the trait is a rare one. Recessives are generally distinguished from dominant mutations by the fact that the latter have an affected parent, and sex linked recessives are distinguished from autosomal recessives by the fact that two normal parents can never have affected female offspring. Familial conditions which will be dealt with in this chapter belong exclusively to the class determined by gene substitutions whose manifestation is not significantly affected by differences of environment such as those to which members of the same fraternity are commonly exposed.

§ 2

The application of quantitative analysis to rare "hereditary" diseases offers no special difficulties and has been discussed already. Factorial analysis of recessive gene substitutions encounters a special difficulty. This is due to the small size of the human family. Rare recessives nearly always arise from matings of the type $H \times H$. In this case neither parent exhibits the trait. Hence unions of the type $H \times H$ which may yield recessive offspring are only distinguishable from unions of the type $H \times D$ or $D \times D$ when there are some recessives among their progeny. If the size of the family is very large, this will always happen. So if human mothers were as prolific as the queen bee, we could simply divide by four the total progeny of all parents

with any offspring of a type unlike themselves. This would give the expected number of offspring on the assumption that the type in question is determined by a single recessive gene substitution. In civilised communities, the average human family is smaller than four. So the mean expectation of recessive offspring is rather less than one when both parents are heterozygous. In a mixed population of such families there will be a certain proportion of families with no recessive offspring and a certain proportion with one, two or more recessive offspring. The number of recessives in a group of families with two parents who are not recessives will therefore be very much higher than a quarter of the total progeny, if we exclude from our census all families with no recessives at all.

If we have a group of families all with parents $H \times H$ and all consisting of, say, four members, it is easily seen that the error is very large. The probability that an offspring of such a union will be recessive is $\frac{1}{4}$, and that it will not be recessive is $\frac{3}{4}$. Families with no recessives, 1 recessive, etc., will occur with frequencies obtained by expanding the binomial $(\frac{3}{4} + \frac{1}{4})^4$, thus:

$$0 \text{ recessives} \quad 4 \text{ normal} \quad (\tfrac{3}{4})^4 = \frac{81}{256}$$

$$1 \text{ recessive} \quad 3 \text{ normal} \quad 4(\tfrac{1}{4})(\tfrac{3}{4})^3 = \frac{108}{256}$$

$$2 \text{ recessives} \quad 2 \text{ normal} \quad 6(\tfrac{1}{4})^2(\tfrac{3}{4})^2 = \frac{54}{256}$$

$$3 \text{ recessives} \quad 1 \text{ normal} \quad 4(\tfrac{1}{4})^3(\tfrac{3}{4}) = \frac{12}{256}$$

$$4 \text{ recessives} \quad 0 \text{ normal} \quad (\tfrac{1}{4})^4 = \frac{1}{256}$$

Thus there will be 81 families with no recessive members in a group of 256 families. We cannot make use of these 81 families with no recessives because they are indistinguishable from families whose parents are $H \times D$ or $D \times D$. Altogether, these 256 families contain 4×256 individuals, of whom $\frac{1}{4}$, or 256, are recessive. If we divided the total number of individuals in the families containing at least 1 recessive by four, we should be led to expect 175 recessives instead of 256. This difficulty can be surmounted in the

following way. An alternative and more recondite method has been proposed by Haldane.*

If q is the expectation of normal and p the expectation of recessive individuals among the offspring of two heterozygous parents, the probability that all the children will be normal is q^s, when there are s members in the fraternity. So for every one s-membered fraternity which *might* contain recessives, q^s will contain no recessives and $1 - q^s$ will contain at least one recessive. Hence, for every fraternity of s individuals with at least one recessive member, the proportion of fraternities which might contain a recessive but in fact do not is

$$\frac{q^s}{1 - q^s}$$

If there are n_s observed fraternities with at least one recessive member, the corresponding number of fraternities which *might* contain recessive offspring is

$$n_s + \frac{n_s q^s}{1 - q^s} = \frac{n_s}{1 - q^s}$$

The total number of individuals corresponding to this is

$$\frac{s \cdot n_s}{1 - q^s}$$

Of these p should be recessive, so that the expectation of recessive offspring in n_s s-membered fraternities of two heterozygous parents with at least one recessive offspring is

$$r_s = \frac{p \cdot s \cdot n_s}{1 - q^s}$$

For a pool of families of varying size ranging from 1 to c, the maximum size of the family, the total number of recessives predicted by genetic theory is

$$\sum_1^c r_s = \sum_1^c \frac{p \cdot s \cdot n_s}{1 - q^s}$$

---

* 1. Hogben (1932), "The Factorial Analysis of Small Families with Parents of Undetermined Genotype," *Journ. Genetics*, Vol. XXVI.

 2. Haldane (1932), "A Method for Investigating Recessive Characters in Man," *ibid.*

For traits determined by a single recessive gene substitution, the values of p and q are respectively $\frac{1}{4}$ and $\frac{3}{4}$ if both parents are heterozygous. The quantity $\dfrac{p \cdot s}{1 - q^s}$ and its variance σ_s^2 can be tabulated, so that it is only necessary to add the products $n_s \times \dfrac{p \cdot s}{1 - q^s}$ or $n_s \sigma_s^2$ for each group of s-membered fraternities. Tables of these functions are given below.

TABLE IX

| Number | (a) $p = \frac{1}{4}$ and $q = \frac{3}{4}$ | | (b) $p = \frac{1}{2} = q$ | |
	$\dfrac{sp}{1 - q^s}$	σ_s^2	$\dfrac{sp}{1 - q^s}$	σ_s^2
1	1·000	0·0000	1·000	0·000
2	1·1428	0·12245	1·333	0·2222
3	1·2973	0·26297	1·715	0·4898
4	1.4628	0·42005	2·134	0·7822
5	1·6389	0·59178	2·581	1·082
6	1·8248	0·77595	3·047	1·379
7	2·0196	0·97024	3·527	1·667
8	2·2225	1·1724	4·015	1·945
9	2·4328	1·3802	4·509	2·215
10	2·649	1·5917	5·005	2·478
11	2·871	1·8053	5·503	2·737
12	3·098	2·0196	6·001	2·992
13	3·329	2·2335	6·5	3·245
14	3·563	2·4464	7·0	3·497
15	3·801	2·6575	7·5	3·748
16	4·041	2·8667	8·0	3·999
17	4·282	3·0738	8·5	4·249
18	4·525	3·2787	9·0	4·500
19	4·770	3·4814	9·5	4·75
20	5·016	3·6821	10·0	5·00

As an illustration of the method of calculation, the following table (Table X) of recorded families of alcaptonuria is given. The standard deviation of the expected number 61·9 is $\pm \sqrt{\Sigma n_s \sigma_s^2} = \sqrt{23 \cdot 5}$, or 4·8 approximately. Thus the discrepancy between the expected and the observed number (66) of alcaptonurics is less than its standard deviation.

* See Appendix II.

TABLE X

ALCAPTONURIA—OFFSPRING OF TWO NORMAL PARENTS

Known Size of Fraternity (s)	n_s	Alcaptonuric Members		$n_s \sigma_s^2$
		Observed	Expected	
1	5	5	5	0·00000
2	8	10	9·14	0·97960
3	5	8	6·48	1·31485
4	2	4	2·93	0·84010
5	3	4	4·92	1·77534
6	3	8	5·47	2·32785
7	2	5	4·04	1·94048
8	3	5	6·67	3·51720
9	1	4	2·43	1·38020
10	1	2	2·65	1·59170
11	3	7	8·61	5·41590
14	1	4	3·56	2·44640
Total ..	37	66	61·90	23·52962

TABLE XI

Size of Family (s)	Number of Families (n_s)	Amaurotics		Variance ($n_s \sigma_s^2$)
		Known	Expected	
2	21	26	24·001	2·5715
3	13	21	16·865	3·4186
4	7	8	10·241	2·9404
5	10	20	16·390	5·9178
6	9	14	16·423	6·9836
7	2	3	4·040	1·8405
8	1	4	2·222	1·1724
9	1	1	2·433	1·3802
10	3	7	7·947	4·7751
11	1	1	2·871	1·8053
12	1	6	3·098	2·0196
Total ..	69	111	106·531	34·825

A second example of close correspondence with the requirements of modern genetic hypothesis is worthy of quoting, because of its clinical interest. Dr. Slome has recently analysed all recorded cases of amaurotic family idiocy, the infantile form of Tay Sach's disease. The disease is lethal, death occurring usually at two years. Hence both parents are necessarily normal. In the table (Table XI) above, all *single* cases are omitted, since they have no statistical significance.

In this case the difference between the observed number of infantile amaurotics (111) and the predicted number ($106 \cdot 5$) is $4 \cdot 5$. The standard deviation is $\pm \sqrt{34 \cdot 8}$, or roughly $5 \cdot 9$. The discrepancy is therefore less than the standard error.

An analogous problem arises when we consider matings of a recessive and a heterozygote. No recessive offspring occur among the offspring of matings $R \times D$. We can only distinguish matings of the type $R \times H$ from $R \times D$ when the progeny include recessive individuals. The only difference between the two cases is that $p = \frac{1}{2}$ instead of $\frac{1}{4}$ and $q = \frac{1}{2}$ instead of $\frac{3}{4}$. In Table IX the values of $\dfrac{p \cdot s}{1 - q^s}$ and of its variance are given for $p = \frac{1}{2} = q$. The ensuing Table XII (opposite) summarises 14 families with albino offspring among marriages between albinos and normal persons, as recorded in Pearson's monograph.

The method of familial analysis outlined in the preceding remarks has been applied to Usher's pedigrees of retinitis pigmentosa (of which there is also a dominant type), xeroderma, juvenile amaurotic idiocy, congenital ichthyosis and total colour blindness. All these conditions have a high incidence of consanguineous parentage like alcaptonuria and albinism. It is not usual to find such a close correspondence between hypothesis and observation as that exemplified in Table XI. There is a simple explanation of why an excess might occur. When a disease is deemed to be familial, cases illustrating a high familial incidence are regarded as medical curiosities. As such they are more likely to be communicated to medical journals. Hence recorded literature is not invariably a random sample of the relevant data. Alcaptonuria is such a rare disease that most cases are reported.* Garrod, one of the earliest writers on the subject, suggested its genetic basis, and drew attention to the need for full family details of all recorded cases. Consequently there is good reason to believe that the sample is fairly representative. Whether the opposite is

* Appendix VI, p. 132.

true of other conditions which do not conform to expectation so well can be tested in a very simple way. The probability that a fraternity will contain more than one affected member obviously increases as the size of the family increases. Hence the discrepancy between prediction and observation on account of biassed sampling will be great for small families and small for large families. For small families biassed sampling will make the observed number of affected individuals greater than it should be, owing to rejection

TABLE XII

ALBINISM—MATINGS OF ALBINOS WITH NORMAL PERSONS

s	n_s	Albinos		$n_s \sigma_s^2$
		Expected	Observed	
1	2	2	2	0
2	2	2·67	3	0·44
3	1	1·71	1	0·49
4	3	6·40	7	2·35
6	2	6·09	8	2·76
7	1	3·52	3	1·67
8	2	8·03	4	3·89
11	1	5·50	4	2·74
Total ..	14	35·92	32	14·33

of fraternities with only one affected member. As Table XIII shows, this can be tested by tabulating the difference (Δ) between expected and predicted numbers of recessives for each size of family and the ratio of this difference to its standard deviation $(\sigma = \sqrt{n_s \sigma_s^2})$.

In the case of total congenital colour blindness, the discrepancy between the expected (63·5) and observed (91) numbers of affected offspring in 37 families with two normal parents is 27·5, or 5·5 times its standard error. If we divide the table horizontally into two halves, it will be seen that the large discrepancies are in the small families. The same is true of juvenile amaurotic idiocy (Sjögren's cases). Haldane has made a second suggestion to account for large excess of affected offspring. This is that parents who have had one or two affected children tend to stop having more. This would also bias

the result in favour of an excess of affected individuals. Both explanations may contribute to disagreement between prediction and observation. In any case, it is important to notice that the agreement is generally satisfactory when small families are excluded.

The algebraical and numerical results which have been deduced for autosomal genes can be applied with very small modifications to sex-linked conditions. The equilibrium proportions in a system

TABLE XIII

MATINGS OF TWO NORMAL PERSONS WITH TOTALLY COLOUR-BLIND
OFFSPRING

Size of Fraternity (s)	Number of Fraternities (n_s)	Number of Affected		Difference Δ $(t_s - r_s)$	σ^2 $(n_s \sigma_s{}^2)$	Δ / σ
		Observed (t_s)	Expected (r_s)			
1	1	1	1·00	0·000	0·000	—
2	2	4	2·28	1·72	0·2449	3·9
3	6	12	7·78	4·22	1·5778	3·3
4	7	18	10·24	7·76	2·9403	4·5
5	5	8	8·19	—0·19	2·9589	0·1
6	3	7	5·47	1·53	2·3278	1·0
7	6	17	12·12	4·88	5·8214	2·0
8	4	13	8·89	4·11	4·6896	1·9
9	2	8	4·86	3·14	2·7604	1·9
10	1	3	2·65	0·35	1·5917	0·3
Total	37	91	63·50	—	24·913	—

of random mating show that recessives will generally occur among offspring of matings between a female carrier and a male who does not exhibit the recessive condition. When the recessive condition is very rare, practically all matings which produce recessives will be of this type, viz.:

$$RD \times DY$$

$$\tfrac{1}{4}RD \qquad \tfrac{1}{4}DD \qquad \tfrac{1}{4}DY \qquad \boxed{\tfrac{1}{4}RY}$$
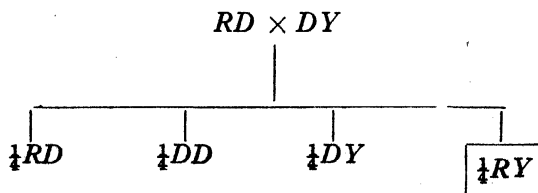
Thus the expectation (p) of recessive offspring for unions of two un-affected parents is a quarter, as for autosomal recessive gene sub-stitution. The only difference is that all affected offspring are males.

If the recessive condition is not very rare, two other types of matings may occur in sufficient frequency for statistical examina-tion. Matings between affected males and females who are carriers may be represented thus:

$$RY \times RD$$

$\frac{1}{4}RR$	$\frac{1}{4}RD$	$\frac{1}{4}DY$	$\frac{1}{4}RY$
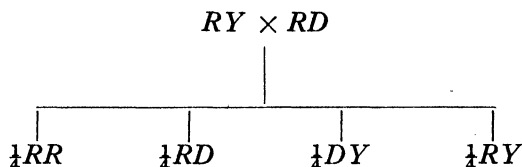
In this case half the offspring are recessives. Such matings can only be distinguished from $RY \times DD$ when there are some recessive offspring. So we must apply the method adopted for matings of the type $R \times H$ when transmission is not sex-linked, and $p = \frac{1}{2} = q$. Matings between an affected female and a normal male should yield *no* affected females and no normal males.

$$RR \times DY$$

$\frac{1}{2}RD$	$\frac{1}{2}RY$
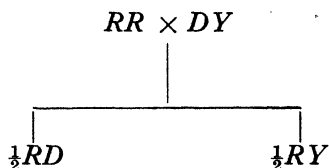
In this case there is no alternative which could lead to confusion. So the method applicable to the analysis of small families with parents of undetermined genotype is not applicable, provided that we pool offspring of all such matings, whether they produce reces-sives or not.

Matings of the type $RY \times RD$ enable us to settle an interesting issue which arises in connexion with haemophilia. The absence of any well-authenticated cases of haemophiliac females has led to controversy. There are three possible explanations. The first is that haemophiliac females have not been found because the disease is very rare. The second is that the gene is lethal in duplicate. The third is that the gene cannot manifest its presence in duplicate owing to the physiological environment of the female soma. By

itself the first explanation would be adequate if there were no well-authenticated matings of the type $RY \times RD$. There are seven families of this type on record. The affected offspring are exclusively male. If the second hypothesis is correct, the sex ratio in such matings should be $1\,♀♀:2\,♂♂$. The data show that 14 out of 25 offspring are normal females and 11 are affected offspring. As far as it goes, the available genetic evidence therefore tends to support the third hypothesis. This receives independent experimental confirmation from the fact that La Fleur Birch claims to have succeeded in remedial treatment of haemophilia by implantation of ovarian tissues and injection of ovarian extracts.

In man, the number of known sex-linked recessive conditions is about the same as the number of autosomal recessives. One class of pedigrees of Leber's hereditary optic atrophy, a large proportion of recorded cases of night blindness, haemophilia, pseudo-hypertrophic muscular paralysis, megalocomea and red-green colour blindness are the best examples. This fact is worthy of comment. Since there are 24 chromosomes in man, there is at first sight no reason why sex-linked recessives should be numerous. It is more remarkable when it is added that there are no very well-established examples of dominant sex-linked conditions in man, and that the number of autosomal dominant mutations provided by medical literature is larger than the number of autosomal recessives. In wild animals, recessives are far more common than dominant mutations. Is man an exception to this rule? The facts stated suggest that he is not. Sex-linked recessive conditions are readily identified by the different incidence of the recessive condition in the two sexes and the absence of any affected females in matings of the more frequent type $(RD \times DY)$. Dominant sex-linked conditions are not so readily recognised as such, because the discrepancy of sex incidence is not conspicuous. In fact, the only simple clue is the absence of affected males in matings of the type $DY \times RR$. On the other hand, autosomal dominants are more easily recognised than recessives on account of their "hereditary" type of transmission. Thus it is not unlikely that the small number of autosomal recessive conditions recognised till now is due to the difficulty of detecting their occurrence. If this is so, there may be a rich harvest for future research directed to the identification of recessive conditions by the systematic study of consanguineous unions.

§ 3

In certain circumstances one type of inheritance is very difficult to distinguish from the transmission of traits determined by a single autosomal recessive gene substitution. This is the complementary action of two independent dominant genes. Parents who exhibit a condition which depends upon the presence of two dominant genes will generally have some affected offspring, so that such traits will be "hereditary" in the clinical sense. The converse is not true. Since one parent may contribute one gene and the other parent its complement, affected offspring may have parents neither of whom exhibit the dominant condition. In that case the trait is "familial" in the clinical sense. Hence mere inspection of a group of clinical pedigrees may lead to the separation of a disease into a dominant and a recessive type when all the data are consistent with a single hypothesis. If the condition is lethal, no individuals will survive to become parents. It will then be classified as a familial disease. If it is very rare, it can still be distinguished from a condition depending upon a single recessive gene substitution, because rare recessives have a high percentage of consanguineous parents. If it is comparatively common, this test is not helpful.

It is therefore necessary to examine the theoretical expectation of affected offspring in matings of various kinds, more especially when neither parent shows the dominant trait. In what follows, the two dominant genes A and B are assumed to be located on different chromosomes and to have frequencies l and m respectively. All the genotypes in a system of random mating would then be as follows:

TABLE XIV

(i) Affected—

$AABB$	l^2m^2
$AaBB$	$2l(1 - l)m^2$
$AaBb$	$4l(1 - l)m(1 - m)$
$AABb$	$2l^2m(1 - m)$

(ii) Not affected—

$AAbb$	$l^2(1 - m)^2$
$Aabb$	$2l(1 - l)(1 - m)^2$
$aaBB$	$(1 - l)^2m^2$
$aaBb$	$2(1 - l)^2m(1 - m)$
$aabb$	$(1 - l)^2(1 - m)^2$

Before proceeding to examine the problem in detail, it is not difficult to see that the transmission of a lethal trait depending upon two independent dominant genes each incapable of producing the manifest trait in the absence of the other may resemble the transmission of a recessive condition more closely than the previous remarks have disclosed. For two unaffected parents to have any affected offspring, one parent must have the gene A without B, and the other must have B without A. If A and B are rare, Aa is much more common than AA, and Bb is much more common than BB. Therefore the bulk of unions between two unaffected parents with affected offspring will be of the type

$$Aabb \times aaBb$$

In this case one-quarter of the offspring will have both A and B. When such a trait is rare, the expectation of affected offspring of two unaffected parents who *may* have affected offspring will tend to a limit which is the same as the expectation for recessive offspring of two parents who are not themselves recessives but may have recessive offspring. It is therefore important to ask how slowly the expectation of affected offspring tends to approach this limit as the incidence of the condition becomes more rare. The interesting feature of the problem is that the expectation approaches its limiting value very closely, even when the condition is not very rare.

Four types of mating between unaffected parents can produce affected offspring. These are as follows:

Mating	Frequency (f)	Proportion of Double Dominants per Family (n)
(a) $aaBB \times AAbb$	$2l^2m^2(1 - l)^2(1 - m)^2$	1
(b) $aaBB \times Aabb$	$4lm^2(1 - l)^3(1 - m)^2$	$\frac{1}{2}$
(c) $AAbb \times aaBb$	$4l^2m(1 - l)^2(1 - m)^3$	$\frac{1}{2}$
(d) $Aabb \times aaBb$	$8lm(1 - l)^3(1 - m)^3$	$\frac{1}{4}$

If $l = m$, i.e. the two dominant genes have the same frequency, the sum of the terms in the middle column is

$$2l^2(1 - l)^4(l - 2)^2$$

The expected number of affected offspring for an s-membered family is

$$\frac{l^2}{(l-2)^2}s + \frac{4l(1-l)}{(l-2)^2} \cdot \frac{\frac{1}{2}s}{1-(\frac{1}{2})^s} + \frac{4(1-l)^2}{(l-2)^2} \cdot \frac{\frac{1}{4}s}{1-(\frac{3}{4})^s}$$

If $l = \dfrac{1}{100} = m$, the familial trait has an incidence of 1 in 2,500.

The *expected* number of offspring does not then differ from the expectation for recessive offspring of two normal parents as given in Table IX by more than 1 per cent. as will be seen below.

TABLE XV

Size of Family	Expected Recessives	Expected Double Dominants ($l = m = 0 \cdot 01$)
2	1·143	1·145
3	1·297	1·301
4	1·463	1·470
5	1·639	1·648
6	1·825	1·837
7	2·020	2·035
8	2·222	2·240
9	2·433	2·454
10	2·649	2·673

For the factorial analysis of familial diseases this result is highly significant. It means that if a condition is determined by two dominant genes of the same order of frequency, it cannot be distinguished from a recessive one, when the incidence in the community is of the order 1 : 2,500. As a matter of fact, the discrepancy is still very small if $l = m = 0 \cdot 02$. A double dominant trait would then be present in roughly one out of every 640 individuals. A recessive trait with a frequency of 1 : 2,500 corresponds to a gene frequency of $0 \cdot 02$. This would raise the frequency of cousin marriages from 1 per cent. to 3 per cent. The latter figure is within the range of variation which occurs in certain sections of the community. So if a familial trait is lethal, like amaurotic family idiocy or

xeroderma, it is only legitimate to regard it as recessive when it is sufficiently rare to justify the expectation of a high excess of consanguineous parentage.

If a familial trait is determined by the interaction of two independent genes whose combined effect is not lethal, there is less difficulty in distinguishing it from a recessive condition, chiefly

TABLE XV

Matings of One Normal and One Affected Parent	Frequency of Mating (f_x)	Population of Affected per Family (n)
(a) $AABB \times aaBB$	$2l^2m^4(1 - l)^2$	1
(b) $AABB \times aaBb$	$4l^2m^3(1 - l)^2(1 - m)$	1
(c) $AABB \times AAbb$	$2l^4m^2(1 - m)^2$	1
(d) $AABB \times Aabb$	$4l^3m^2(1 - l)(1 - m)^2$	1
(e) $AABB \times aabb$	$2l^2m^2(1 - l)^2(1 - m)^2$	1
(f) $AaBb \times aaBB$	$8lm^3(1 - l)^3(1 - m)$	$\frac{1}{2}$
(g) $AaBb \times aaBb$	$16lm^2(1 - l)^3(1 - m)^2$	$\frac{3}{8}$
(h) $AaBb \times AAbb$	$8l^3m(1 - l)(1 - m)^3$	$\frac{1}{2}$
(i) $AaBb \times Aabb$	$16l^2m(1 - l)^2(1 - m)^3$	$\frac{3}{8}$
(j) $AaBb \times aabb$	$8lm(1 - l)^3(1 - m)^3$	$\frac{1}{4}$
(k) $AaBB \times aaBB$	$4lm^4(1 - l)^3$	$\frac{1}{2}$
(l) $AaBB \times aaBb$	$8lm^3(1 - l)^3(1 - m)$	$\frac{1}{2}$
(m) $AaBB \times AAbb$	$4l^3m^2(1 - l)(1 - m)^2$	1
(n) $AaBB \times Aabb$	$8l^2m^2(1 - l)^2(1 - m)^2$	$\frac{3}{4}$
(o) $AaBB \times aabb$	$4lm^2(1 - l)^3(1 - m)^2$	$\frac{1}{2}$
(p) $AABb \times aaBB$	$4l^2m^3(1 - l)^2(1 - m)$	1
(q) $AABb \times aaBb$	$8l^2m^2(1 - l)^2(1 - m)^2$	$\frac{3}{4}$
(r) $AABb \times AAbb$	$4l^4m(1 - m)^3$	$\frac{1}{2}$
(s) $AABb \times Aabb$	$8l^3m(1 - l)(1 - m)^3$	$\frac{1}{2}$
(t) $AABb \times aabb$	$4l^2m(1 - l)^2(1 - m)^3$	$\frac{1}{2}$

because matings between an affected parent and a parent who does not exhibit the trait will nearly always have some affected offspring. This is especially true when the trait is relatively common. However, this criterion must be used with circumspection, because the ratio of matings of the type $R \times H$ to matings of the type $R \times D$ is often unexpectedly high even when a recessive trait is rare, like albinism. Matings of one affected and one normal parent when two dominant genes determine a familial trait are given in the table shown above.

Further calculation may be simplified by collecting the terms with the same expectation per family.

$$m_1 = \overset{e}{\Sigma}f_x + f_m + f_p$$
$$m_2 = \overset{a}{f_n} + f_q$$
$$m_3 = f_f + f_h + f_k + f_l + f_o + \overset{r}{\Sigma}f_x$$
$$m_4 = f_g + f_i \quad .$$
$$m_5 = f_j$$

The expected proportion of affected offspring in all fraternities with only one affected parent is then

$$\frac{m_1 + \tfrac{3}{4}m_2 + \tfrac{1}{2}m_3 + \tfrac{3}{8}m_4 + \tfrac{1}{4}m_5}{\overset{t}{\underset{a}{\Sigma}f_x}}$$

This approaches a lower limit of $\tfrac{1}{4}$ as the trait beomes more rare. For equal frequencies of the two dominant genes, it has a value of $0 \cdot 30$ when the incidence of the trait is 4 per cent, and $0 \cdot 28$ when the incidence of the trait is 1 per cent. There is therefore a simple test for distinguishing such conditions from traits determined by a single recessive gene substitution. The proportion of affected offspring of *all* matings involving one affected and one normal parent should lie between 1 and $0 \cdot 25$. Generally speaking, a much lower proportion would be obtained by taking all matings involving only one affected parent when a trait depends on a single recessive gene substitution, because matings of the type $R \times D$ yield no affected offspring and are far more numerous than matings of the type $R \times H$.

There remains the possibility of distinguishing between a trait determined by two dominant genes and what are sometimes loosely called "incompletely" dominant characters in the sense that a single dominant gene may not manifest its presence in the simplex condition unless certain conditions—not always available—are present in its environment. If a double dominant condition is comparatively rare, the expectation of affected in offspring in matings of two affected parents may without serious error be taken to be

$$\Sigma t_s = \Sigma \tfrac{1}{4} \cdot \frac{s \cdot n_s}{1 - (\tfrac{3}{4})^s}$$

Also the expectation of affected offspring for all unions involving only one affected parent may be taken to be $\frac{1}{4}$, and for all unions of two affected parents $\frac{9}{16}$. The recorded data for the type of dwarfism known as ateleiosis conform closely to the first two requirements but not the last.*

◢ § 4

It has already been pointed out that analysis of this kind is only applicable to gene substitutions whose manifestation is not affected by differences of environment such as commonly occur within the framework of the family unit. In applying it to diseases which do not appear till comparatively late in life, sibs who die young must not be included in the total size of the fraternity. The success of these methods depends upon the way in which the data are collected. They are more or less applicable in the measure that we are entitled to look upon recorded data as a representative sample of the distribution of a trait in different families of the population from which they are taken. Much of the available evidence in medical literature is not representative in this sense.

One type of bias from which recorded data may suffer is illustrated by a recent experience of the writer. A consultant, distinguished for his special knowledge of a fairly common disease, sent him sixty pedigrees in which it was present. An analysis of these pedigrees yielded remarkably consistent correspondence with the hypothesis that the disease depends upon the interaction of two independent dominant genes. Though this result was highly gratifying, it was necessary to ascertain how the data had been collected. It transpired that the physician had written to about three hundred of his patients, asking them if they knew of any relatives who were also sufferers. He then investigated in detail the family histories of those patients who replied in the affirmative. Presumably, a large proportion of his patients were isolated cases with no affected relatives. Had these been included in the sample, they could not have failed to introduce large discrepancies, unless they were all cases of only children with no offspring of their own.

Another source of bias has been suggested already. This is the

* Hogben (1932), "The Genetic Analysis of Familial Traits. II: Double Gene Substitutions," *Journ. Genetics*, Vol. XXV.

tendency to place on record fraternities illustrating a high familial incidence of a rare complaint, omitting those which do not. In concluding this chapter it may be helpful to specify some general directions which should be observed in the collection of pedigrees if the fullest use is to be made of them.

(i) If the physician is devoting himself to the study of a single disease or a small group of diseases, he must treat all cases as if they had equal value for genetical purposes. That is to say, he should not restrict his records of family histories to patients with relatives who are also sufferers from the same complaint. It is just as important for the geneticist to know about the patient who has no affected relatives as to know about the patient who can boast of a proud lineage of fellow sufferers.

(ii) If the physician obtains the family history of any isolated case of a rare familial disease, he should place it on record, whether there are affected sibs or not. If he only records cases which illustrate the occurrence of the disease in several members of a fraternity, the geneticist will find that the familial incidence is much higher than the correct hypothesis demands.

(iii) It is always important to give the age at death of both parents, the age of the mother at the birth of each of her offspring, the correct birth order of all sibs, including miscarriages, sex and age at death of all sibs, and age of onset of the disease of all affected members in a pedigree. The ages of death and onset are not so important in the case of congenital complaints which are recognised at birth. They are very important when a disease is not congenital, or, if it is congenital, when it is not recognised till the child begins to grow up. The reason for this has been given. If a disease does not appear till after puberty, no individuals who die before puberty should be counted in determining the familial incidence. The age of the mother at the birth of the offspring and the correct birth rank of each sib is important, because it may draw attention to an important class of environmental factors which limit the manifestation of a gene difference. This is well illustrated by "mongols," and possibly some other classes of amentia, when the uterine environment is an important factor in the manifestation of a trait.

It may be asked whether such precise knowledge of the genetic mechanism of disease has any practical consequence. The answer is that it provides the physician with an exact answer to a question which actually arises in his practice. Two healthy parents who have

given birth to a child who is an amaurotic family idiot or a haemophiliac may wish to know what likelihood there is that they will do so again, if they have another child. The physician can tell them that the odds are exactly one to three in favour of each subsequent child being afflicted in the same way. There are even odds that any child will carry the defective gene and transmit it to future generations. The odds are one to three that any subsequent child will neither have the disease nor carry the gene for it.

SUMMARY

When a condition arises from a single recessive gene substitution, it is recognisable by its high familial incidence in fraternities of which both parents are unaffected. To test the statistical agreement between the observed familial incidence and Mendelian hypothesis it is necessary to take into account the fact that a high proportion of families in which no recessives actually occur are necessarily excluded from the census. For certain rare conditions, notably alcaptonuria, amaurotic family idiocy, congenital ichthyosis and xeroderma, there is remarkably close correspondence between the recorded familial incidence and what would be predicted by factorial analysis.

When a condition is determined by two dominant genes located on different chromosome pairs, the expectation of affected offspring in affected fraternities with two unaffected parents does not differ significantly from that predicted by the factorial hypothesis for single recessive gene substitutions. So if the condition is lethal, the two cases are indistiguishable without data concerning consanguineous parentage.

Owing to the small size of the human families, the expected values for the familial incidence diverge very slowly as the number of genes which are involved increases.*

* See Appendix III.

THE INTERDEPENDENCE OF NATURE AND NURTURE

§ I

In the first chapter reference was made to the fact that some writers of human genetics have used the method of correlation to give an estimate of the relative importance of nature and nurture as sources of variability in human populations. We shall now discuss how far this procedure is justifiable. When Mendel's law was first applied to animals by Bateson and Punnett, the use of correlation methods became the focus of a lively controversy in which there was misunderstanding on both sides. Correlations between relatives calculated on the assumption that the effects of genes are perfectly additive, i.e. that no dominance exists, were found to agree fairly well with observed data for metrical characters such as stature. This led to a somewhat barren discussion, because the experimental school were at first inclined to regard dominance as an essential feature of hereditary transmission. When dominance is complete, the correlations between relatives are not independent of the gene frequencies and will in general be lower than when the measurement of the heterozygote is exactly intermediate between the two pure homozygous types. Among experimental biologists the belief which gained ground was that correlation coefficients are nothing more than descriptive devices to exhibit the resemblance of relatives, yielding no information concerning whether such resemblance is due to the distribution of genes or the distribution of differences in the external environment. In his memoir entitled "The Correlation between Relatives on the Supposition of Mendelian Inheritance,"* R. A. Fisher put forward a contrary view in 1918. Fisher attempted a synthesis between the particulate theory of inheritance and the problem of nature and nurture as it had been formulated by Galton and his successors. In the domain of animal genetics, Fisher's conclusions were of importance at the time in so far as they bear on the universality of Mendel's first law. Since they were stated, the application of Mendelian concepts to the study of metrical characters

* Fisher (1918), "The Correlation between Relatives on the Supposition of Mendelian Inheritance," *Trans. Roy. Soc. Edin.*, Vol. LII.

has been tested by alternative methods, described by East and Jones in their book *Inbreeding and Outbreeding*. For human genetics the question has an ulterior significance, because of the peculiar difficulty of distinguishing between differences due to environment and differences due to heredity in human populations and because of the far-reaching sociological implications of this distinction.

Fisher's contribution to the discussion was noteworthy both for the thoroughness with which he assailed the mathematical intricacies of a purely genetical theory of correlation and for the particular conclusions about nature and nurture advanced in his memoir. His objective was to determine the respective contributions which nature and nurture make to the variability of a normal population, using the mean square deviation as the measure of variability. In his own words, it was "to ascribe to the constituent causes" (heredity and environment) "fractions or percentages of the total variance which they together produce." In his final summary, the most important statement is as follows:

By means of the fraternal correlation it is possible to ascertain the dominance ratio and so distinguish dominance from all non-genetic causes such as environment, which might tend to lower the correlations. . . . The fact that this excess of the fraternal correlation is very generally observed is itself evidence in favour of the hypothesis of cumulative factors. On this hypothesis it is possible to calculate the numerical influence not only of dominance but of the total genetic and non-genetic causes of variability. An examination of the best available figures for human measurements shows that there is little or no indication of non-genetic causes. The closest scrutiny is invited on this point, not only on account of the practical importance of the predominant influence of natural inheritance, but because the significance of the fraternal coefficient in this connection has not previously been realised.

Since these words were written, a substantial body of new data concerning metrical characters of human beings has accumulated, notably in connexion with the study of twin resemblance. On this account, Fisher's theory is entitled to closer scrutiny than has been undertaken hitherto. The importance of Fisher's hypothesis of cumulative factors is beyond dispute. What merits special consideration is the attempt to draw up a balance sheet of nature and nurture. An attempt will here be made to examine the way in which subsequent progress in the measurement of differences due to environment and differences due to genes, as such differences can be studied in the laboratory, bears upon the discussion of such differences in human populations. The first question to ask is whether

the technique of correlation can be used to detect the existence of differences due to environment and differences due to heredity.

To this question the answer is certainly in the affirmative. Table XVII (pages 94–95) shows the results of recent observations on the intelligence quotients of twins set forth in the form of correlation coefficients. There is general agreement concerning the conclusion that monozygotic twins are intellectually more alike than dizygotic twins of like or of unlike sex.* In other words,· the intellectual resemblance between individuals who are known to be genetically alike is greater than the intellectual resemblance between individuals who may be presumed to be genetically different. Few biologists would hesitate to draw the conclusion that intellectual differences may arise because of gene differences. In so far as it is true that the correlation between ordinary sibs is lower than for fraternal twins, we are equally entitled to surmise that intellectual differences may arise within the family, because of differences of environment associated with birth rank. An alternative possibility should not be overlooked. Haldane and Crew (1925) found that crossing over in poultry varies with the age of the parent. This might tend to increase the variance of sibs born at different times as compared with fraternal twins.

Another example of the way in which the technique of correlation may be used to *detect* the existence of genetic differences is provided by an elementary application of the theory of inbreeding. Like homogamy (*vide infra*), inbreeding will tend to raise the correlation between sibs. Hence the fraternal correlation for a group of individuals whose parents are first-cousins should be somewhat higher than the fraternal correlation of a group of individuals whose parents are unrelated, provided both groups are taken at random from the same population. This method has not yet been used. It merits special consideration because it is free from an objection to the study of twins as a means of detecting géne differences affecting intellectual characteristics. On account of their greater physical resemblance, identical twins may tend to work together, play together and be exposed to the same sources of infection to a greater extent than fraternal twins. There is, therefore, an element of legitimate doubt concerning the correct interpretation of the greater intellectual

* Stocks (*Ann. Eugen.*, Vol. V., 1933) on the other hand finds that the similarity of fraternal twins tested with the Terman group test is not less than that of identical twins.

NATURE AND NURTURE

TABLE XVII

TWIN CORRELATIONS OF IQ

(The number in brackets indicate the numbers of pairs examined)

Author	Test	Like Sex	
		Identical Pairs	Fraternal Pairs
Merriman (1924) ..	Stanford-Binet—		
	5–9 years	—	—
	10–16 years	—	—
	Beta—		
	5–9 years	—	—
	10–16 years	—	—
	National Intelligence—		
	5–9 years .. .:	—	—
	10–16 years	—	—
Lauterbach (1925) ..	Terman	—	—
Wingfield (1928) ..	Average of various ..	0·90 ± 0·02 (45)	—
Holzinger (1929) ..	Binet	0·88 ± 0·02 (50)	0·63 ± 0·06 (52)
	Otis	0·92 ± 0·01 (50)	0·62 ± 0·06 (52)
Stocks (1933)	Terman	0·84 ± 0·03 (78)	0·87 ± 0·02 (66)
Herrman and Hogben (1933)	Otis	0·86 ± 0·04 (65)	0·49 ± 0·08 (96)

resemblance of the former. A third use of the technique of correlation to detect gene differences will be made clear at a later stage. Genetic theory allows us to predict what correlations of first-cousins would arise in a homogeneous environment, when transmission is sex-linked. With respect to the sex of individual members of a pair of cousins and the sex of their related parents, first-cousins may be classified in ten groups, for which the correlation coefficients calculated on this assumption differ widely. A corresponding order of size in the observed values would be sufficient justification for concluding that sex-linked genes affect the character measured.

So long as the use of correlation methods is confined to the recognition that gene differences or differences due to nurture exist, there is little room for disagreement. The difficulties of interpretation begin when we attempt to clarify what is meant by calculating "the

TABLE XVII—*continued*

TWIN CORRELATIONS OF IQ

(The numbers in brackets indicate the numbers of pairs examined)

Like Sex		All-Like Sex Pairs	Unlike Sex Pairs	All Fraternal (Like Sex and Unlike Sex)
Boy Pairs	Girl Pairs			
0·80±0·08 (10)	0·91±0·03 (19)	0·88±0·03 (19)	0·77±0·06 (18)	—
0·89±0·03 (17)	0·81±0·05 (21)	0·86±0·03 (38)	0·30±0·14 (20)	—
0·93±0·05 (7)	0·71±0·10 (19)	0·92±0·03 (16)	0·52±0·15 (12)	—
0·75±0·08 (13)	0·89±0·03 (16)	0·84±0·04 (29)	0·64±0·09 (19)	—
0·92±0·04 (7)	0·96±0·01 (24)	0·96±0·01 (31)	0·75±0·07 (23)	—
0·89±0·03 (14)	0·92±0·02 (37)	0·86±0·02 (61)	0·83±0·04 (28)	—
0·81±0·03 (71)	0·73±0·04 (63)	—	0·56±0·06 (78)	—
—	—	0·82±0·03 (76)	0·59±0·08 (26)	0·70±0·04 (57)
—	—	—	—	—
—	—	0·86±0·02	0·38±0·1 (65)	—
0·69±0·04 (126)	0·63±0·05 (141)	0·66±0·02 (267)	0·53±0·06 (138)	—

numerical influence . . . of the total genetic and non-genetic cause, of variability." In his illuminating essay on the *Notion of Causes* Bertrand Russell has pointed out that few words are used with greater ambiguity in scientific discussion. The biometrical treatment of variability inherited from Galton a tradition of discourse in which the ambiguity of the concept of causation completely obscured the basic relativity of nature and nurture. Since then this relativity has become increasingly recognised through experiments involving the use of inbred stocks in physiological laboratories, especially in connexion with experimental work on diet. It is therefore necessary to examine with great care what we mean when we make measurements of a genetic difference and a difference due to environment.

The diagram on page 96 will avoid any danger of undue

abstraction in the definitions we shall adopt. It is based on Krafka's*
data. In the fruit fly *Drosophila* there is a series of mutations charac-
terised by extensive reduction of the number of facets in the com-
pound eye. Two such mutations are designated "low-bar" and
"ultra-bar." The actual number of facets varies with the temperature
of the environment in which the larvae develop. In the diagram
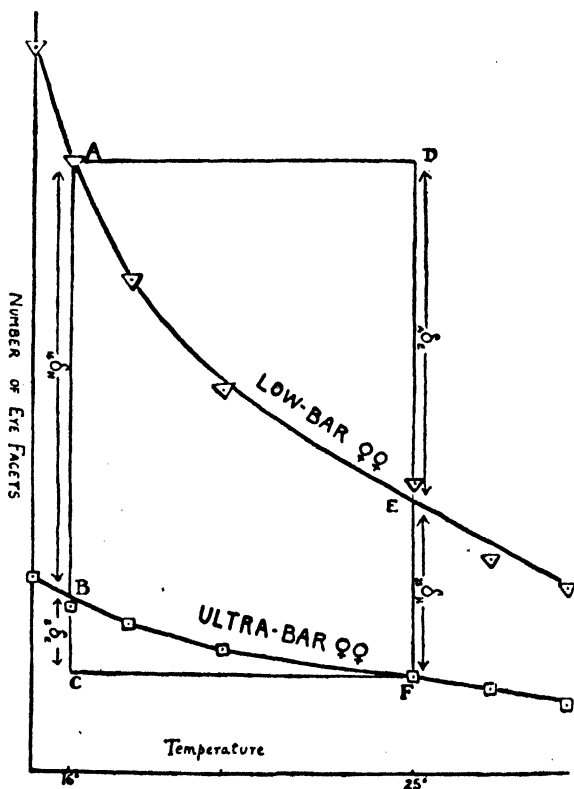
FIG. 2

(Fig. 2), the distance AB, measured along the ordinate ABC,
represents the difference $(_{16}\delta_H)$ of the two stocks, both cultured at
$16°$ C. The distance $EF(_{25}\delta_H)$ represents the difference between the
two stocks, both cultured at $25°$ C. The length $BC(_B\delta_E)$ represents
the difference between the measurement of ultra-bar individuals
cultured respectively at $16°$ and $25°$ C. The length $DE(_A\delta_E)$ represents

* Krafka (1920), *Journ. Gen. Physiol.*, Vol. 2.

the difference between the facet number of low-bar individual cultured at 16° and 25° C. respectively. $EF(_{25}\delta_H)$ and $AB(_{16}\delta_H)$ each correspond to what the experimental biologist means by a genetic difference. $BC(_B\delta_E)$ and $DE(_A\delta_E)$ each respresents what the experimental biologist calls a difference due to environment.

Clearly we are on safe ground when we speak of genetic difference between two groups measured in one and the same environment or in speaking of a difference due to environment when identical stocks are measured under different conditions of development. Are we on equally safe ground when we speak of the contribution of heredity and environment to the measurements of genetically different individuals or groups measured in different kinds of environment? Suppose we measure a low-bar stock kept continuously at 16° C. and an ultra-bar stock kept continuously at 25° C. The observed difference will be represented by the length AC or DF. How much of AC or DF is due to heredity and how much to environment? The question is easily seen to be devoid of a definite meaning. We might be tempted to say that the genetic contribution is the difference which would exist if both stocks had been cultured at the same temperature. This could be done in an infinite number of ways. If they were both cultured at 16° C., heredity would contribute the difference AB. We might also be tempted to say that the contribution of environment represents what the difference would be if all the flies belonged to the same stock. Obviously this can be done in at least two ways. Keeping the same difference of environment we might substitute low-bar individuals for the ultra-bar stock at 25° C. The difference between the two sections of the population would then be represented by DE. If we substituted ultra-bar individuals for the low-bar stock at 16° C., the difference would be BC. Either ED or BC is equally entitled to be regarded as the contribution of environment.

The literature of experimental physiology is not wanting in examples of such divergent curves representing the relation between the measurement of a character and the strength of the environment. There is no reason to multiply instances in order to show the need for extreme care in formulating the problem of nature and nurture in quantitative terms. At this stage it is sufficient to recognise what we mean by correlation as a measure of resemblance between relatives in a homogeneous environment and what we mean by

correlation as a measure of resemblance between genetically identical individuals when classified in some specified relationship to the external conditions of development. These will form the subject of preliminary examination. The legitimacy of using the technique of correlation to ascribe to heredity and environment "fractions or percentages of the total variance which they together produce" will be reserved for discussion at a later stage.

§ 2

A purely genetic theory of correlation is evidently valid, so long as it is concerned with gene differences whose manifest effect is not affected by differences of environment to which different members of a population are exposed. With this implicit restriction, Pearson and Yuel were the first writers to show how correlation coefficients for relatives might be deduced on the basis of Mendel's theory. Fisher undertook the difficult task of removing any limitations concerning the number of genes, the degree of dominance or epistacy, independent segeregation or the absence of assortative mating. A full discussion of all the issues raised involves some difficult and extremely laborious mathematics. The basic assumptions underlying the main conclusions which have been drawn from such studies can be illustrated by studying the simplest possible types of correlation between relatives. This does not require any statistical knowledge except the way in which a coefficient of correlation is calculated and a knowledge of the genotypic ratios for a poulation in equilibrium, when mating occurs at random. If a is the frequency of a recessive gene and b of its dominant allelomorph, the equilibrium state for an autosomal gene substitution is

$$a^2(R): 2ab(H): b^2(D)$$

For sex-linked genes, when the sex ratio is unity, it is

$$\tfrac{1}{2}a(RY): \tfrac{1}{2}b(DY): \tfrac{1}{2}a^2(RR): ab(RD): \tfrac{1}{2}b^2(DD)$$

The frequency of different types of mating has already been given on pages 39 and 41.

The simplest possible assumptions on which a correlation coefficient for relatives can be determined are (a) that mating occurs at random, (b) that only one gene substitution affects the character measured, (c) that the environment is homogeneous. Two cases will

suffice to show how a correlation coefficient may be deduced on these assumptions. The correlation coefficient for brothers when transmission is sex-linked is the easiest to deal with. All possible pairs of brothers may be set out in tabular form as below:

Frequency	Parents	Male Offspring	Pairs of Brothers		
			RR	RD	DD
a^3	$RY \times RR$	R	1	0	0
$2a^2b$	$RY \times RD$	$\frac{1}{2}R + \frac{1}{2}D$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
ab^2	$RY \times DD$	D	0	0	1
a^2b	$DY \times RR$	R	1	0	0
$2ab^2$	$DY \times RD$	$\frac{1}{2}R + \frac{1}{2}D$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
b^3	$DY \times DD$	D	0	0	0

Adding the numbers in each column on the right-hand side appropriately weighted by those in the column on the extreme left, we find the frequency of recessive pairs (RR) is

$$a^3 + \tfrac{1}{2}a^2b + a^2b + \tfrac{1}{2}ab^2 = a^2 + \tfrac{1}{2}ab$$

Proceeding in this way we obtain a contingency table:

	R	D
R	$a^2 + \tfrac{1}{2}ab$	$\tfrac{1}{2}ab$
D	$\tfrac{1}{2}ab$	$b^2 + \tfrac{1}{2}ab$

To determine the correlation coefficient we have to assign some numerical value to the difference between R and D. If we make the difference between the measurement of the recessive and dominant genotype in the standard environment the unit of measurement, we can put $R = 0$ and $D = 1$. We therefore find* that $r = 0 \cdot 50$.

This case is simplified by the absence of a heterozygous class. Thus the degree of dominance does not affect the result which is

* See Appendix IV.

independent of the frequency (a) of the recessive gene. When there is a heterozygous class to be taken into consideration, the value of r is affected by how much H differs from D. If its measurement is exactly intermediate between R and D, the result is independent of the rarity of· the recessive gene. If it approximates more or less closely to D, this is not so. As an illustration, the correlation of parent and child when transmission is autosomal will be given. The possible pairs are tabulated below:

Frequency of Mating	Parents	Offspring	Pairs of Genotypes taking Parent with Child					
			RR	RH	RD	HH	HD	DD
a^4	RR	R	1	0	0	0	0	0
$4a^2b$	RH	$\frac{1}{2}R + \frac{1}{2}D$	$\frac{1}{4}$	$\frac{1}{2}$	0	$\frac{1}{4}$	0	0
$2a^2b^2$	RD	H	0	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0
$4a^2b^2$	HH	$R + \frac{1}{4}H + \frac{1}{4}D$	0	$\frac{1}{4}$	0	$\frac{1}{2}$	$\frac{1}{4}$	0
$4ab$	HD	$\frac{1}{2}H + \frac{1}{2}D$	0	0	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
b^4	DD	D	0	0	0	0	0	1

Adding the columns as before, we find:

$$RR = a^4 + a^3b = a^3$$
$$HH = 2a^2b^2 + a^3b + ab^3 = ab \ . \ . \ ., \text{ etc.}$$

This yields a contingency table as follows:

	R	H	D
R	a^3	a^2b	0
H	a^2b	ab	ab^2
D	0	ab^2	b^3

To calculate a correlation coefficient from such a table it is necessary to assign some value to H. If we assume that the heterozygote differs from the measurement of the two homozygous genotypes by an equal amount, we can calculate it by putting $R = 0$, $H = \frac{1}{2}$,

$D = 1$. In this case a's and b's cancel out and we are left with an arithmetical solution $r = 0 \cdot 5$, independent of the rarity of the recessive genotype. If dominance is complete, we can put $R = 0$, $H = D = 1$. We thus get a fourfold table:

	R	$(H + D)$
R	a^3	a^2b
$(H + D)$	a^2b	$b(1 + ab)$

The value of r is then found to be

$$1 + a$$

The result therefore depends upon the frequency of the recessive genotype. If a quarter of the population are recessives (i.e. $a = \frac{1}{2}$), $r = 0 \cdot 33$. A more general solution is obtained by representing the measurement of the heterozygous class by a fraction A, which may be given any value between $\frac{1}{2}$ and 1. It is then found that the value of r is only independent of the frequency of the recessive genotype

TABLE XVIII
AUTOSOMAL GENE SUBSTITUTIONS

	Heterozygote Intermediate $(A = \frac{1}{2})$	Dominance Complete $(a = \frac{1}{2})$
Individual and		
(a) Ancestral line:		
Parent	0·5000	0·3333
Grandparent	0·2500	0·1667
Great-grandparent	0·1250	0·0833
Great-great-grandparent ..	0·0625	0·0417
(b) Collaterals:		
Sibs	0·5000	0·4167
Uncle or aunt	0·2500	0·1667
First-cousins	0·1250	0·0833
Second-cousins	0·03125	0·0208

TABLE XIX*

Some Correlation Coefficients for Relatives on the Supposition of Sex-linked Transmission

	Heterozygote Intermediate $(A = \frac{1}{2})$	Dominance Complete $(a = \frac{1}{2})$
A. Parental		
Father-son	0·00	0·00
Mother-daughter	0·50	0·33
Father-daughter	0·71	0·58
Mother-son	0·71	0·58
B. Fraternal		
Sister-sister	0·75	0·67
Brother-brother	0·50	0·50
Mixed sibs	0·33	0·20
C. First-Cousins		
(i) Maternal (parents' sisters):		
Girl pairs	0·187	0·125
Boy pairs	0·375	0·375
Mixed pairs	0·250	0·133
(ii) Paternal (parents' brothers):		
Girls pairs	0·250	0·167
Boy pairs	0·000	0·000
Mixed pairs	0·000	0·000
(iii) "Mixed" (parents' brother and sister):		
Girl pairs	0·125	0·083
Boy pairs	0·000	0·000
†Mixed pairs (a)	0·000	0·000
Mixed pairs (b)	0·167	0·067

when the heterozygote is exactly intermediate in its measurement between the two homozygous genotypes.

The above two tables (XVIII and XIX) show the values of the correlation coefficient calculated on the same assumptions for relatives of different kinds. It will be noticed that the value of the

* Hogben (1932), "The Correlation of Relatives on the Supposition of Sex-linked Transmission," *Journ. Genetics*, Vol. XXVI.

† (a) Mother of girl is sister to father of boy. (b) Father of girl is brother to mother of boy.

fraternal correlation coefficient when transmission is autosomal is the same as the parental correlation coefficient when the heterozygote is intermediate in the sense already defined. When dominance is complete and $\cdot a = \frac{1}{2}$, the parental is reduced by twice as much as the fraternal coefficient. By increasing the number of gene substitutions which affect the measurement of a variable characteristic, we can increase the number of compartments in a contingency table such as those which have been illustrated in the foregoing remarks to as many as we choose. The more general treatment for n gene substitutions has been given by Fisher. Fisher, has shown that the parental deviation with respect to dominance is twice the fraternal for all degrees of dominance, independently of the number of genes and of their frequencies. When the heterozygote is intermediate, the values given in the accompanying table are the same, whether a

TABLE XX

	Eye Colour	Joint of Little Finger	Stature	Hair Colour	Cephalic Index	Width of Wrist
Parent-child	0·49	—	0·51	—	—	—
Brothers	0·51	—	0·53	—	0·49	—
First-cousins	0·43	0·34	—	0·29	—	0·29

single gene substitution or a large number of genes affect the measurement of differences between individuals.

The figures given in the first table agree in some cases with observed correlations of relatives for measurable human characteristics, a few of which are given in the table above. They are taken from communications by Pearson and Lee and by Elderton. It will be seen that the values obtained for measurements of human relatives are generally higher than the preceding treatment permits. This is specially true of first-cousins. So we are led to ask, what are the several agencies which affect the value of a coefficient of correlation? Four categories merit special consideration. These are (a) dominance, (b) the contribution of sex-linked genes, (c) assortative mating, (d) a heterogeneous environment.

The significance of dominance calls for cautious interpretation. Dominance lowers the theoretical correlation for parent and child more than for sibs. Table XX illustrates the fact that observed parent-child correlations are typically lower than correlations for sibs. Of

itself this does not tell us very much. The environment of father and son is less homogeneous than that of two brothers, a fact which sometimes gives rise to friction. Indeed, a great deal of the world's dramatic literature deals with the way in which the social environment of parents unfits them for the privilege of exporting advice to foreign generations. The data published by Pearson and Lee do not give any clear indications of the influence of sex-linked genes upon the correlation of relatives. For stature, the correlation of father and son is o·514, which is actually higher than for mother and son, o·494, and for eye colour the coefficient for sister and brother is o·553, which is higher than for sister and sister, o·537. This is not surprising. There is no reason to expect that sex-linked genes play a largé part in determining variability, since there are twenty-three pairs of human chromosomes alike in both sexes and only one pair unequally matched.

The effect of assortive mating in the sense that like tends to choose like (homogamy) is to raise correlations between relatives. A general treatment of this problem is laborious. There is no difficulty in illustrating the truth of the statement by the extreme example, when only like genotypes mate with one another. If transmission is autosomal, the contingency table for husband and wife is then:

<div align="center">

Husband

		R	H	D
		R	α	o	o
Wife	H	o	β	o
	D	o	o	γ

</div>

In this case the correlation between husband and wife is unity instead of zero, as it would be if mating occurred at random. The fraternal contingency table is easily constructed. There are a fraternities consisting of R only, γ of D only, and β of R, H and D in the proportions $\frac{1}{4}:\frac{1}{2}:\frac{1}{4}$. All possible pairs of genotypes in fraternities of the last class occur in the following proportions:

$$\tfrac{1}{16}RR + \tfrac{1}{4}RH + \tfrac{1}{8}RD + \tfrac{1}{4}HH + \tfrac{1}{4}HD + \tfrac{1}{16}DD$$

So the contingency table for brothers is:

	R	H	D
R	$\alpha + \tfrac{1}{16}\beta$	$\tfrac{1}{8}\beta$	$\tfrac{1}{16}\beta$
H	$\tfrac{1}{8}\beta$	$\tfrac{1}{4}\beta$	$\tfrac{1}{8}\beta$
D	$\tfrac{1}{16}\beta$	$\tfrac{1}{8}\beta$	$\gamma + \tfrac{1}{16}\beta$

The value of r when the heterozygote is intermediate and the proportion of all three genotypes is the same is $0 \cdot 80$. Correlations between husband and wife approaching unity are not recorded for any physical trait. For stature Pearson gives the value $0 \cdot 28$ and for eye colour $0 \cdot 10$. It is not clear how far such coefficients measure departures from random mating in the purely genetical sense, since they are not based on spouses taken from a homogeneous mating population.

§ 3

When we speak of a homogeneous environment in the present context, strictly physical homogeneity is not implied. Every gene difference requires some special condition of the environment to make it manifest to the investigator. It happens that the appropriate conditions in which some gene differences can be recognised are present in almost any environment in which the fertilised egg will develop. The environment is then *neutral* in Professor Levy's sense, and such gene differences belong to the class of "real isolates." Having outlined the genetical theory of correlation in its more elementary aspects under this restriction, let us now examine how correlations may arise in virtue of the distribution of environmental variables. Considerable mathematical skill and patience have been devoted to the construction of a purely genetical theory of correlation. Hitherto the complexity of the influence which differences of environment may exercise, has received less attention, though the study of twin resemblance has lately shown the urgent need for a careful examination of the problem. The issues involved will be more readily seen if we pause at this point to examine the form of the correlation coefficient.

If we have a series of paired measurements $x_1y_1, x_2y_2 \ldots x_ny_n$, we commonly calculate the correlation between x and y by the product moment formula:

$$r_{xy} = \frac{\Sigma xy}{N\sigma_x\sigma_y}$$

The mean variance* for each pair is

$$\frac{\Sigma(x-y)^2}{2N} = V_p$$

If V is the variance of the whole population of x's and y's, a simple transformation* gives

$$1 - r = \frac{V_p}{V}$$

When the correlation coefficient is exhibited in this form, it is evident that r is greater or less according as differences between different values of x *or* different values of y taken at random are large compared with differences between corresponding values of x *and* y, *whatever the nature of the agency which is responsible for such differences.* This means that differences of environment such as tend to increase the difference between two members of the same pair affect the correlation coefficient in the opposite sense to such differences of environment as tend to make members belong to different pairs more different.

The bearing of this statement upon the study of variability in a human population is important because differences of environment which affect the measurement of a metrical character are not necessarily distributed with uniformity. Differences to which members of the same family or different families living at one and the same social level are exposed may be very much less than differences to which individuals belonging to families taken from different social levels are exposed. Experiment shows that ultra-violet light has a considerable influence upon growth in mammals. In Great Britain some families live continuously in the sooty atmosphere of an industrial area. Others spend their winters on the Riviera. That the study of fraternal correlations leads some students of human inheritance to the conclusion that there is little or no indication of "non-genetic

* See Appendix V.

causes" tending to produce differences of stature in human populations may throw more light upon the limitations of statistical technique and their method of interpretation than upon the physiology of human growth. A human society may be crudely compared to a badly managed laboratory in which there are many cages each containing a pair of rats and their offspring. The rats are of different breeds. The cages are at different distances from the window. Different cages receive different rations. Rats in the same cage cannot all get to the feeding trough together. So some get more meal than others. The cage corresponds to the family as a unit of environment. The rats in each cage constitute the family as a genetic unit.

To the human geneticist the practical and theoretical importance of the environment is so great that it is justified to dwell upon the issue involved in the preceding remarks. A hypothetical illustration suggested by one of the earlier researches of Hopkins upon diet will assist to make it clear. Imagine a series of cages each containing a pair of rats from the same homozygous stock, each cage supplied with a basic ration of *zein*, to which is added a different quantity of trytophane. An experiment might be carried out so that the range of measurements for, say, body-weight is discontinuous for the pairs. The paired measurements then form a series proportional to the trytophane added to the basic ration. A correlation table drawn up for the body-weight of pairs sharing the same cage would yield a value of r equal to unity. Provided the stock is homogeneous, the result will be the same if each pair of rats is a pair of sibs. The two cases would not necessarily be the same if the stock were genetically heterogeneous. If the rats in the same cage are unrelated and genetically different, and if all the cages receive the same amount of zein and trytophane, the correlation coefficient for body-weight will be zero. What will happen if the rats sharing the same cage are sibs of a mixed stock, when the ration for all cages is the same? Let us assume that the quantity of tryptophane added to the basic ration is so chosen that gene differences with respect to susceptibility to tryptophane content can manifest themselves. If we neglect the contribution of sex-linked genes and dominance deviations, the correlation will be 0.5. This figure measures the tendency of two sibs to resemble one another when reared in an effectively homogeneous environment.

Using the same heterogeneous stock, let us return to the original

procedure. Each cage receives a different quantity of tryptophane. The effect of genetic heterogeneity is to increase the deviations both of the x's and the y's. Whether we measure r for cage-pairs which are sibs or cage-pairs which are unrelated individuals, the value of r will be less than when a homogeneous stock is used. For the case when the pairs are sibs, r must lie between $0 \cdot 5$ and $1 \cdot 0$. The experiment may be varied in other ways. We might divide each cage by a partitition, distribute the food at random so that two members of a pair ordinarily receive a different quantity of tryptophane and determine the value of r for body-weight of the cage-pairs as before. If the stock is homogeneous, there will always be zero correlation whether cage-pairs are sibs or unrelated. If the stock is heterogenous, there will be zero correlation for unrelated cage-pairs. For sibs, r would lie between $0 \cdot 5$ and zero. We might combine the two procedures, varying the food given to the two rats on either side of a partition while making the difference between the ration given to two members of a cage-pair small in comparison with the difference between the mean quantity given to one cage-pair and the mean quantity given to another cage-pair. The result would now depend upon the order of smallness. If the stock is homogeneous, the value of r will be greater than zero and will be the same whether the cage-pairs are sibs or unrelated. If the stock is heterogenous, the value of r will always be less when the members of a cage-pair are unrelated individuals than it would be if they were sibs. In the latter case unaided common sense does not make it obvious whether the value of r will be greater or less than $0 \cdot 5$.

The effect of increasing the difference in environment as between one pair and another is, therefore, to increase the coefficient of correlation above the value theoretically deduced on the assumption that the environment is homogeneous. Increasing the differences of environment to which members of the same pair are exposed tends to lower it. That an observed correlation between brothers very closely coincides with what can be deduced from a purely genetic theory of correlation does not justify the inference that the observed differences between individuals mainly belong to the class of genetic differences which manifest themselves in any environment consistent with existence. The way in which we select relatives for the determination of a correlation coefficient may be such that the effect of differences due to environment between one pair and another balance the differences between two individuals of

the same pair. Hence it is not difficult to see why Sewall Wright*
obtained higher correlations for litter mates than for sibs of different
broods in certain measurements on guinea-pigs. It is not surprising
that some workers find a higher correlation for intelligence of
dizygotic twins than for the intelligence of sibs born at different
times, or that Freeman, Holzinger and Mitchell,† in their work
on the intelligence quotient, found that the correlation between
foster-sibs is of the same order as the correlation of sibs adopted
into different homes at birth.

Referring to the formula which connects the coefficient of correla-
tion with the mean variance within each fraternity and the variance
of the population as a whole, Fisher makes the following comment:
"For stature the coefficient of correlation between brothers is about
0·54, which we may interpret by saying that 54 per cent of their
variance is accounted for by *ancestry* alone and that 46 per cent
must have some other explanation. It is not sufficient to ascribe this
last residue to the effects of environment. . . . The simplest hypothesis
is . . . that the large variance among children of the same parents
is due to the segregation of those factors in respect of which the
parents are heterozygous." This statement occurs at the outset.
The balance sheet of nature and nurture which follows later stands
or falls by the meaning which we attach to the word *ancestry*. Its
use in the writings of statistical geneticists invites the closest scrutiny.
If differences of nurture were distributed uniformly within the
family unit and between one family unit and another, the concept
of ancestry would involve no ambiguity in human genetics. In the
laboratory we can culture stocks of oviparous animals, arranging the
conditions so as to ensure that any slight differences to which
different individuals are exposed are as likely to involve two related
individuals as to involve two unrelated individuals. Then and then
only are we safe, when we speak of "the random external effects of
environment" and deal with nature and nurture as independent
variables. In fact this condition is not strictly realised when we are
studying a viviparous animal. A further complication arises when
we are dealing with social species like the primates, which live in
family groups.

* Sewall Wright (1926), *American Naturalist*, Vol. 60.
† Freeman, Holzinger and Mitchell (1928), "The Influence of Environment
on the Intelligence, School Achievement and Conduct of Foster-Children,"
27th *Year-book N.S.S.E.*, Part I, Chap. 9.

When a species is viviparous, progeny of a given mother share common features of pre-natal environment either because they belong to the same litter or, if they do not, because they are housed in the uterus of an individual with a given genetic constitution. When a species is social, offspring of the same parents generally grow up in an environment which is more homogeneous than the common environment of two unrelated individuals. In, the case of human beings this is true as regards both physical culture (food, sunlight, sleep and exercise) and social tradition. The familial constellation of environmental variables is not confined to a single fraternity of human beings. Generally the environment in which an individual is reared is more like the environment in which his parent was reared than the environment of an unrelated person of the same age as his parent. Hence the ancestry of an individual, that is to say what he shares with or derives from his ancestors, includes: (i) a system of genetic relations, (ii) a system of developmental relationships determined by the uterine environment but correlated with the preceding, and (iii) a framework of social and physical environment also related to the genetic "ancestral" relationship.

In the table (Table XXI) on pages 112–13, an attempt has been made to show how different elements of the social complex are implicitly included in specifying the ancestry of individuals, when a human population is made the subject of investigation by correlation technique. The several components are subdivided as follows:

(1) *Genetic constitution.*—This involves both the genotype and the dominance deviations. Fisher's values for autosomal transmission are those given. They represent the result which would be found if the environment were perfectly homogeneous. The table neglects the contribution of sex-linked genes which would involve specifying every type of sex relationship separately.

(2) *Uterine environment.*—We have to distinguish between individuals which develop in the same uterus at the same time (++), individuals which develop in the same uterus at different times (+o) and individuals which develop in different wombs (oo).

(3) *Family environment.*—Members of a pair may be reared together (+) or reared apart (o). If they are reared together, they may have the same (+) or different (o) birth rank. They may be of the same sex (+) or opposite sex (o). A third environmental variable is more subtle and, like sexual similarity, involves a correlation

between genetic constitution and the degree of association between the environments to which two individuals are exposed. Thus two identical twins tend to associate more than two sibs or fraternal twins whose different equipment of genes may lead them into different kinds of environment. If one is weakly and the other strong, they may go to different schools or be pampered to a greater or less extent. Because they have different temperaments, they may choose different companions. In all such ways genetic differences may be associated with differences of environment acting as a limiting factor to the exhibition of other genes which two such individuals share.

(4) In the two columns on the extreme right of Table XXI differences between individuals belonging to different pairs are tabulated. It has already been pointed out that the results obtained from correlation studies may be affected in opposite ways by increasing the differences of environment to which both members of a given pair are subjected, and increasing differences of environment to which members of different pairs are subjected. The specification of the environmental framework therefore involves a statement of the kind of environmental differences to which members of different pairs are subjected as well as the kind of environmental differences to which members of the same pair are exposed. If a fraternal coefficient of correlation is based exclusively upon the sons of dairy-farmers, it does not follow that it will be exactly the same as if it were based on a population made up of pairs of sons taken from widely different social groups.

The data on which Fisher's analysis of variability with respect to stature is actually based are those collected from a comparatively homogeneous social group by Pearson and Lee. It would not be expected that mathematical analysis could extract from them conclusions concerning variability in a population of different social levels. Using the term ancestry in a purely genetic sense, Fisher arrives at the conclusion that: "It is very unlikely that so much as 5 per cent of the total variance is due to causes not heriditable." The assumption on which this estimate is based is that the contribution of nurture can be evaluated from the discrepancy between observed correlations for relatives and correlations calculated from purely genetic considerations. The foregoing illustration shows that this is not a legitimate assumption. In the laboratory, we could take a genetically homogeneous stock of rats and obtain a

TABLE XXI

	Members of Same Pair		
	I. Genetic		II. Uterine Environment
	Genotypic Correlation	Phenotypic Correlation	
(a) Identical twins reared together	1·0	1·0	$++$
(b) Identical twins reared apart from birth	1·0	1·0	$++$
(c) Fraternal like-sex twins reared together .. .. .. }	$\frac{1}{2}$	$\frac{1}{2} \cdot \dfrac{\sigma^2 - \frac{1}{2}e^2}{\sigma^2}$	$++$
(d) Fraternal like-sex twins reared apart from birth	$\frac{1}{2}$	same as above	$++$
(e) Mixed twins reared together	$\frac{1}{2}$	same as above	$++$
(f) Mixed twins reared apart from birth	$\frac{1}{2}$	same as above	$++$
(g) Like-sex sibs reared together	$\frac{1}{2}$	same as above	$+\circ$
(h) Like-sex sibs reared apart from birth	$\frac{1}{2}$	same as above	$+\circ$
(i) Like-sex sibs (all pairs) reared in institution	$\frac{1}{2}$	same as above	$+\circ$
(j) Mixed sibs reared together ..	$\frac{1}{2}$	same as above	$+\circ$
(k) Mixed sibs reared apart ..	$\frac{1}{2}$	same as above	$+\circ$
(l) Mixed sibs (all pairs) reared in institution .. .. }	$\frac{1}{2}$	same as above	$+\circ$
(m) Father-child	$\frac{1}{2}$	$\frac{1}{2} \cdot \dfrac{\sigma^2 - e^2}{\sigma^2}$	$\circ$
(n) Mother-child	$\frac{1}{2}$	$\frac{1}{2} \cdot \dfrac{\sigma^2 - e^2}{\sigma^2}$	●
(o) First-cousins	$\frac{1}{8}$	$\frac{1}{8} \cdot \dfrac{\sigma^2 - e^2}{\sigma^2}$	∞
(p) Unrelated individuals reared together	$\circ$	$\circ$	∞

TABLE XXI—*continued*

Members of Same Pair					Members of Different Pairs	
III. Familial Environment				IV. Social Level	Same Family Group	Same Social Level
Same Family	Same Birth Rank	Sex	Choice			
+ +	+	+	+ + +	+	o	+ or o
o	—	-+	+	+ or o	o	+ or o
+ +	+	+	+ +o	+	o	+ or o
o	—	+	+	+ or o	o	+ or o
+ +	+	o	+ +o	+	o	+ or o
o	—	+	+	+ or o	o	+ or o
+ +	o	+	+ +o	+	o	+ or o
o	—	+	+	+ or o	o	+ or o
+ +	—	+	+	+	+	+
+ +	o	o	+	+	o	+ or o
o	—	o	+	+ or o	o	+ or o
+ +	o	o	+o	+	+	+
+	o	o	—	+ or o	o	+ or o
+	—	—	—	+ or o	o	+ or o
o	—	—	—	+ or o	o	+ or o
+ +	o	+ or o	?	+ or o	o	+ or o

fraternal correlation in complete agreement with what a purely genetical theory of correlation demands by the expedient of adjusting the differences of environment to which offspring of the same parents and individuals belonging to different sibships are respectively exposed. In such a situation the contribution of gene differences to the observed variance would actually be zero.

Fisher justifies the construction of a balance sheet for the contributions of nature and nurture by the fact that variances are additive when the several "causes" of variability are independent. This implies that each genotype has an equal chance of experiencing —with their respective probabilities—each of the available kinds of environment. Such a state of affairs may have some resemblance to an orphan asylum. It has little relevance to human populations in general. The viviparous habit of the human species, the existence of the family as a social institution, the stratification of human society in widely different social levels all conspire to create a framework of environment which is intimately related to the distribution of genetic differences.

§4

A much abused philosopher of the nineteenth century has remarked that "all the mysteries which seduce speculative thought into mysticism find their solution in human practice and in concepts of this practice." In such terms the experimental biologist will seek for any significance in a balance sheet of nature and nurture. The only practical significance which Fisher's analysis of variability seems to admit is that, if it were correct, we could only reduce variance with respect to stature in a human population by 5 per cent or less if the environment were perfectly uniform. From an experimental standpoint, what do we mean by making the environment uniform? We can do so in an infinite number of ways, some tending to bring out genetic differences which were not previously measurable, others tending to obscure genetic differences which were measurable before. Has a balance sheet of nature and nurture any meaning in this sense, unless we assume that the variance of a population, if affected at all, is necessarily diminished when the environment is made more homogeneous? Such an assumption is certainly false, as a single example suffices to show.

In the fruit fly *Drosophila*, the mutant stock, designated "abnormal

abdomen" owing to its characteristic deformity when grown in a dry culture, is not distinguishable from the wild type if the larvae are grown on moist food. Imagine a large laboratory with many bottles of culture media, some dry and some moist, providing food for a mixed stock of fruit flies, a small proportion of which belong to the mutant strain with the gene for vestigial abdomen. Keeping the stock the same, we might make the environment more homogeneous in one of two ways, either making all the bottles dry or all the bottles moist. If we make all the bottles dry, the mutant gene will be incapable of manifesting its presence. Variability will be diminished with respect to the difference under consideration. If we make all the bottles moist, a larger proportion of larvae with the mutant gene will hatch out as flies with the mutant deformity. That is to say, there will be an increase in variability. This example is not an isolated one, and it is specially relevant to human biology. The effect of extending to all classes of society the educational opportunities available to a small section of it would presumably be that of increasing variability with respect to educational attainment. The effect of depriving the more favoured of their special educational advantages would be to diminish variability in educational attainment. Either policy would involve the elimination of environmental differences.

There will be even less room for misunderstanding if we examine a metrical situation concerning which we have definite experimental knowledge. In Krafka's experiments on the effect of temperature upon the number of ocelli in two bar-eye mutants of *Drosophila* (Fig. 2), the following values were obtained for females at 15° and 25° C.:

	"Low-bar"	"Ultra-bar"
15 C. 15 C.	198 74	52 25

The simplest kind of population in which uncorrelated differences*

---

* It makes no difference to the practical interpretation of the balance sheet whether we take correlated or uncorrelated differences. The former is simpler. For instance, all the low-bar individuals might be cultured at 15° C. and all the ultra-bar at 25° C. The mean would then be 107 and the

due to environment and genetic differences coexist is one composed of only two genotypes, each having equal access to either of two kinds of environment. We might take a population half composed of low-bar and half composed of ultra-bar stock, equal numbers of each kind being exposed to temperatures of 15° and 25° C. The mean number of eye facets for the entire population would be 85 and the variance 3,906. We can eliminate differences due to nurture by keeping both stocks at 15° C. The mean would then be 120·5 and the variance would be 4,692 to the nearest integer. We can eliminate all differences due to environment by keeping both stocks at 25° C. The mean would then be 49·5 and the variance 600. Have either of these estimates any special priority as a measure of the contribution of heredity alone to the observed variance? We can eliminate all differences due to heredity by substituting an equal number of low-bar females for that half of the population made up of ultra-bar stock. The mean is now 131·5 and the variance is 3,306. We could also eliminate all differences due to heredity by substituting ultra-bar stock for the low-bar individuals. The mean would then be 38·5 and the variance would be 182. Which of these two estimates give the contribution of environment alone?

Are we to understand that the "standard environment" which must be defined to give any meaning to a balance sheet of nature and nurture is a statistical average? True, such an average would have a definite significance in the foregoing example. It would not have an equally definite significance for the study of human populations exposed to an indefinitely large number of as yet unmeasurable and unidentified environmental differences. Nor would any balance sheet drawn up on such an assumption entitle us to set limits to changes which could be produced by controlling the environment.

In whatever sense Fisher himself intended his balance sheet to be interpreted, there is no doubt that many writers on human biology entertain the belief that biometrical estimates of this kind do entitle us to set such limits. On the basis of such statements as the previous quotation about stature, it is often argued that the results of legislation directed to a more equitable distribution of medical care must be small, and that in consequence we must look to selection for any noteworthy improvement in a population.

variance 6,664. We still obtain different estimates according to the way in which we eliminate differences of environment and differences due to heredity.

This is rather like saying that the difference between black and white is negligible because an inkpot thrown into a tank of china clay has very little effect on the latter. The gross nature of the fallacy is easily seen with the help of a parable. Imagine a city after a prolonged siege or blockade extending over a number of years. The available supplies of food containing the necessary vitamins have long since been exhausted in the open market. Young children still growing are stunted in consequence and weigh on the average 20 per cent less than pre-war children. One biochemist has a small stock of crystallised vitamins which he has reserved for his family of four, who grow up normally. There are, let us say, a million stunted children to four healthy ones. A party of rabid environmentalists is clamouring for peace. The Government appoint an official inquiry of statisticians. They report that far less than 1 per cent of the observed variance with respect to body-weight is due to differences in diet, that the improvement produced by change in diet if peace were made would therefore be negligible, and that eugenic selection would solve the problem of how to keep a community alive without vitamins if the war could be prolonged for a few more millennia.

It requires no subtlety to see what is wrong with this conclusion. If only four in a million and four children had sufficient vitamins for normal growth, the effect of differences in the vitamin content of the diet to the observed variance in the population would be statistically a negligible quantity. In spite of this, the mean body-weight of the population could be increased by 30 per cent if all children received a ration with a vitamin content equivalent to the greatest amount available to any child in the same population. No appreciable deterioration would result from depreciating the ration of the four most favoured to the same level of vitamin content as those of the million less favoured.

This does not mean that the technique of correlation cannot be used to draw attention to differences which are determined by nurture. It is one thing to detect the existence of gene differences or differences due to differences of environment. It is another thing to "calculate the numerical influence of the total genetic and non-genetic causes of variability." In discussing the framework of social environment in a population we have to reckon with larger social units as well as with the family and the uterus as units of environment. A community can be divided into social levels such that families belonging to one social level are more alike with respect to

hygiene, diet, access to sunlight and culture than families belonging to different social levels. The proportions of families at different social levels vary considerably. Where there exist large inequalities of wealth, samples taken at random from the community tend to be overweighted by families at one social level and underweighted by families at another. Environmental differences which distinguish one social level from another tend to be obscured for another reason. Administrative difficulties make it much easier to obtain large samples of families living at one social level than at another. By confining an inquiry to a group of elementary-school children in the London area, the difficulty of dealing with a large number of separate authorities is eliminated. Special methods of inquiry must be devised if we are concerned with the rôle of significant differences of environment to which individuals belonging to families at different social levels are exposed.

§ 5

There is one situation in which it is possible to ascribe some measure of singularity to what we mean by eliminating all genetic differences. This arises in connection with the study of twin resemblance, when we compare variability among a population composed of pairs of fraternal twins of like sex taken from different families and pairs of identical twins taken from different families. We are then comparing the variability *within the family* when gene differences and differences due to family environment acting on individuals of the same birth rank and sex are both present with the variability within the family when such genetic differences as distinguish one member of a family from another member of the same family do not enter into the result. In this case each family may be regarded as a partial fraternity of two. The genotype ratios for the group as a whole are the same in both cases, but there are no genotypic differences within the fraternity itself. Variability in a group of such partial fraternities may be expressed in terms of the mean difference or the mean squared difference. Holzinger* has proposed two formulae for the nature nurture "ratio" respectively based on one or the other measure of variability.

* Holzinger (1929), "The Relative Effect of Nature and Nurture Influences on Twin Differences," *Journ. Educ. Psych.*, Vol. XX, No. 4.

The first formula is written:

$$T = \frac{\Delta_f - \Delta_i}{\Delta_i}$$

Here Δ_f is the mean difference between dizygotic twins of like sex and Δ_i is the mean difference between identical twins; the denominator measures the mean difference when there are no genetic differences within the family, the numerator measures the difference between the variability which exists when both sources of variability are present and the variability when gene differences have been eliminated in the sense defined above. The values of T given by different authors who have investigated the intelligence quotient of twins are as follows:

Author	T	Test
Holzinger	0·5–1·0	Binet and Otis
Tallmann	0·61–0·62	Stanford Revision
Herrman and Hogben ..	0·92–0·62	Otis

The second formula given by Holzinger is as follows:

$$t^2 = \frac{r_i - r_f}{1 - r_i}$$

In this case r_i and r_f represent the correlation coefficients for identical and for fraternal twins respectively. The meaning of this formula is analogous to the preceding one except in so far as the variance or mean square differences take the place of mean differences. Using the relation defined by the difference formula (page 106), we can put:

$$1 - r_i = \frac{V_i}{V}$$

and

$$1 - r_f = \frac{V_f}{V}$$

In these expressions V_i and V_f represent the mean variance of the partial fraternities respectively composed of identical and fraternal

twins and V is the variance of the population from which they are taken. We may thus rewrite Holzinger's formula:

$$t^2 = \frac{V_f - V_i}{V_i}$$

For variations in the intelligence quotient, estimates of t^2 have been given by two groups of workers as follows:

        Holzinger   ..   ..   .. $2 \cdot 0 - 3 \cdot 8$
        Herrman   ..   ..   .. $2 \cdot 3 - 2 \cdot 6$

Holzinger's first formula is misleading, if taken to signify that the numerator measures the average difference due to heredity alone because variability ascribed to the influence of diverse agencies is not additive, when measured in mean differences. The second is open to another criticism. To the present writer it does not seem that the separate sources of variability can be treated as truly additive even when we have eliminated differences between the environment of one family and another in this way, unless we interpret the significance of gene differences in a very special sense. In dealing with intellectual resemblances, we have to remember that the greater physical similarity of identical twins may be associated with a more restricted choice of environment. The fact that t^2 is greater than unity might signify that genetic differences are more important than differences of environment in producing striking discrepancies between members of the same family and the same birth rank. There would be no reason to doubt this conclusion, if it were not for the fact that striking differences of environment are more likely to occur between two non-identical twins than between two identical twins.

It has been pointed out to the writer by Professor H. J. Muller that, when there is sufficient available material to determine the value of $_a r_i$ the correlation for identical twins reared apart, it will be possible to extend the conception of the *nature-nurture ratio* beyond the limits of the family unit. Assuming that the correlation for pairs of unrelated individuals taken at random from the same population is zero, Holzinger's second formula interpreted in this more general sense becomes:

$$\frac{_a r_i}{1 - {_a r_i}}$$

The values obtained might differ greatly according to the magnitude of the differences of environment in the population and their distribution. It would be interesting to compare the values obtained for such a formula with those derived from Holzinger's second formula in its original form. At the same time it would be necessary to exercise great caution in the interpretation of its meaning and sociological application. It might encourage us to hope for conspicuous results from selective breeding. It could not justify us in discouraging attempts to control the social environment.

SUMMARY

Three main conclusions emerge from the foregoing discussion.

The technique of correlation can be used to draw attention to the existence of genetic differences of or differences due to environment, provided the selection of data is appropriate to the kind of differences we wish to detect.

The belief that a comparison between observed correlations of relatives and correlations based upon purely genetical assumptions provides us with a measure of the influence of nurture is not justified, because of the close relationship between the distribution of gene differences and differences due to environment in populations of viviparous animals which live in families, especially when, as with human populations, the environment of different families may differ greatly.

In so far as a balance sheet of nature and nurture has any intelligible significance, it does not entitle us to set limits to changes which might be brought about by regulating the environment.

The application of statistical technique in the study of human inheritance is beset with pitfalls. On the one hand the experimental difficulties of the subject-matter necessitate recourse to mathematical refinements which can be dispensed with·in animal breeding. On the other there is the danger of concealing assumptions which have no factual basis behind an impressive façade of flawless algebra. The student may recall the words of Wilhelm Ostwald:

Among scientific articles there are to be found not a few wherein the logic and mathematics are faultless but which are for all that worthless, because the assumptions and hypotheses upon which the faultless logic and mathematics rest do not correspond to actuality. .

APPENDIX I

Using the symbols employed throughout this book, the standard deviations of the Mendelian ratios calculated on the basis of random mating as in Table I, Chapter II (page 45), is given by Snyder as follows:

(i) For the proportion $\left(\dfrac{a}{1+a}\right)^2$ *Taster* $\times$ *Taster*,

$$\sigma^2 = \frac{a^2(1-a)}{N(1+a)^5}$$

(ii) For the proportion $\left(\dfrac{a}{1+a}\right)$ *Taster* $\times$ *Non-taster*,

$$\sigma^2 = \frac{1-a}{4N(1+a)^3}$$

In these expressions, N is the total number of individuals tested and a is the frequency of the recessive gene, which is taken as the square root of the observed proportion of non-tasters in the sample.

The standard deviation of the incomplete binomial given in tabular form in Table IX, Chapter IV (page 76), is deduced as follows:

If a sibship contains s members and has parents $(H \times H)$ who may have affected offspring, the frequencies of families with 0, 1, 2 . . ., r recessives form the binomial series of which the general term is ${}^sC_r p^r q^{s-r}$. Proceeding by the usual method for determining the mean and standard deviation for the complete binomial, we may make a table of frequencies (f) and occurrences (r). The mean is thus given by

$$\frac{\Sigma r(f)}{\Sigma(f)},$$

and the mean square deviation referred to zero occurrences as origin is

$$\frac{\Sigma r^2(f)}{\Sigma(f)}.$$

If all families with no recessives are excluded, $\Sigma(f) = 1 - q^s$. Hence the mean (M) is

$$\frac{\overset{s}{\underset{1}{\Sigma}} r^s C_r p^r q^{s-r}}{1 - q^s}.$$

The standard deviation referred to the mean as origin is given by

$$\sigma_s{}^2 = \frac{\overset{s}{\underset{1}{\Sigma}} r^2 \cdot {}^s C_r p^r q^{s-r}}{1 - q^s} - \left[\frac{\overset{s}{\underset{1}{\Sigma}} r^s C_r p^r q^{s-r}}{1 - q^s}\right]^2 \quad . \quad . \quad . \quad . \quad (1)$$

The reduction may be effected thus:

$$\frac{\partial}{\partial p} \cdot \overset{s}{\underset{1}{\Sigma}} {}^s C_r p^r q^{s-r} = \frac{1}{p} \overset{s}{\underset{1}{\Sigma}} r^s C_r p^r q^{s-r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\frac{\partial^2}{\partial p^2} \cdot \overset{s}{\underset{1}{\Sigma}} {}^s C_r p^r q^{s-r} = \frac{1}{p^2} \overset{s}{\underset{1}{\Sigma}} r^{2s} C_r p^r q^{s-r} - \frac{1}{p^2} \overset{s}{\underset{1}{\Sigma}} r^s C_r p^r q^{s-r}. \quad . \quad . \quad (3)$$

Also:

$$\frac{\partial}{\partial p}\overset{s}{\underset{1}{\Sigma}}{}^sC_r p^r q^{s-r} = \frac{\partial}{\partial p}[(p + q)^s - q^s]$$

$$= s(p + q)^{s-1} = s, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$\frac{\partial^2}{\partial p^2}\overset{s}{\underset{1}{\Sigma}}{}^sC_r p^r q^{s-r} = \frac{\partial}{\partial p}[s(p + q)^{s-1}]$$

$$= s(s - 1). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Combining (2) and (4) we have:

$$\overset{s}{\underset{1}{\Sigma}}r\,{}^sC_r p^r q^{s-r} = sp \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Combining (3) and (5):

$$\overset{s}{\underset{1}{\Sigma}}r^2\,{}^sC_r p^r q^{s-r} = s(s - 1)p^2 + sp. \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Substituting in (5) and (6) from (9) and (10):

$$M = \frac{p \cdot s}{1 - p^s} = p \cdot c_s \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Thus:

$$\sigma_s{}^2 = \frac{s(s - 1)p^2 + sp}{1 - q^s} - \frac{s^2 p^2}{[1 - q^s]^2}$$

$$= \frac{s(1 - q)}{[1 - q^s]^2}\{sq^s(q - 1) + q(1 - q^s)\}$$

Putting $\sigma_s{}^2 = k_s$, the standard deviation of $n_s p c_s$ is

$$\sqrt{\overline{k_s + k_s \ldots \text{to } n_s \text{ terms}}} = \sqrt{\overline{n_s k_s}}. \quad . \quad . \quad . \quad (9)$$

Similarly, since there is zero correlation between successive samples of families of different sizes, the total expectation of recessives in a pool of families of various sizes is

$$\Sigma n_s p c_s \pm \sqrt{\overline{\Sigma n_s k_s}}. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

The analysis outlined in § 3, Chapter IV, raises the question: can we identify traits determined by two independent recessive genes? The first point to notice is that unless such traits are exceedingly rare we cannot expect a significant excess of consanguineous unions among parents of double recessives. For two recessive genes with frequencies a_1 and a_2, the formula corresponding to that given on page 59 approaches a limit of

$$\frac{x}{x + 256 a_1 a_2}$$

When a double recessive condition is *extremely* rare, the majority of double recessives will have parents

$$AaBb \times AaBb.$$

One-sixteenth of the offspring of such unions will be double recessives (*aabb*). In that case p tends to a lower limit of $\frac{1}{16}$ instead of $\frac{1}{4}$ for single recessive conditions. At this lower limit, the expected number of affected persons in an s-membered fraternity with two parents who are not themselves affected but have at least one affected child is

$$\frac{1}{16} \cdot \frac{s}{1 - \left(\frac{15}{16}\right)^s}$$

Even at this limit the expectation does not differ by a very large amount from the expectation calculated on the basis $p = \frac{1}{4}$.

Recently Sjögren (*Acta Psychiatrica et Neurologica*) has published a noteworthy case history of 34 interconnected families from a small inbred community in Norway, where there has been for some years a higher incidence of low-grade amentia. The individuals classified as affected were unable to learn reading, writing or counting, being arrested at a mental age of about six years or less. Altogether 50 affected persons occurred in these 34 families, and in 13 fraternities more than one affected sib is recorded. If this type of low-grade amentia were determined by a single recessive gene, the expected number would be 65. If it were determined by two recessive genes located in different chromosomes, the

expected number would be 40 at infinite dilution of the recessive genes. Even on this assumption the agreement would be better in the latter case. As a matter of fact, the gene frequencies are not low for this group. So matings of the types $Aabb \times aaBb$, etc., would be comparatively common, tending to raise the value of p from $\frac{1}{16}$ towards $\frac{1}{4}$. For the gene frequencies consistent with Sjögren's data, the expectation would be about half-way between 40 and 65. Thus Sjögren's data as they stand are more easily reconciled with a multiple than with a single gene hypothesis. Against this it may be pointed out that the single gene hypothesis advocated by Sjögren himself is consistent with his data on one (or both) of two subsidiary assumptions: (a) that the low-grade type of amentia in this community is a somewhat lethal condition; (b) that its manifestation depends upon special conditions of uterine environment. The first is a likely supposition, and the second is supported by Pearson's data in *The Handicapping of the First-Born*.

If the number of genes which determine a condition is greater than three, there will be more variables than observational equations for testing how many there are. The expected proportion of recessive offspring of two unaffected parents tends very slowly to its limiting value at infinite dilution, and the small size of the human family makes the divergence very small for expected recessives over a wide range of the number of recessive gene substitutions. If the frequency of all the recessive genes is the same $(a_1 = a_2 = a_3 \ldots = a_r)$, the expected number of recessives in families of *four* with two normal parents having at least one recessive offspring is as follows:

Incidence of Familial Trait in the Population p	Number of Genes				
	1	2	3	4	5
1 : 1,000	1·463	1·320	1·176	1·160	1·152
1 : 10,000	1·463	1·172	1·120	1·104	1·096
1 : 1,000,000	1·463	1·144	1·084	1·096	1·060

These values are calculated on the assumption that mating occurs at random. They show that the method of familial analysis must be used with great circumspection in the absence of evidence concerning consanguineous parentage.

We proceed with an algebraic correlation table of this type, thus:

	0	$\frac{1}{2}$	1
0	a	β	γ
$\frac{1}{2}$	β	δ	η
1	γ	η	θ

If M_x is the mean of all the x's and M_y the mean of all the y's

$$M_x = M_y = \tfrac{1}{2}(\beta + \delta) + (\gamma + \theta) + \tfrac{3}{2}\eta = M.$$

The working is as follows:

δ_x	δ_y	$\delta_x\delta_y$	f	$\delta_x{}^2 = \delta_y{}^2$
M	M	M^2	a	M^2
M	$M - \frac{1}{2}$	$M^2 - \frac{1}{2}M$	β	M^2
M	$M - 1$	$M^2 - M$	γ	M^2
$M - \frac{1}{2}$	M	$M^2 - \frac{1}{2}M$	β	$M^2 - M + \frac{1}{4}$
$M - \frac{1}{2}$	$M - \frac{1}{2}$	$M^2 - M + \frac{1}{4}$	δ	$M^2 - M + \frac{1}{4}$
$M - \frac{1}{2}$	$M - 1$	$M^2 - \frac{3}{2}M + \frac{1}{2}$	η	$M^2 - M + \frac{1}{4}$
$M - 1$	M	$M^2 - M$	γ	$M^2 - 2M + 1$
$M - 1$	$M - \frac{1}{2}$	$M^2 - \frac{3}{2}M + \frac{1}{2}$	η	$M^2 - 2M + 1$
$M - 1$	$M - 1$	$M^2 - 2M + 1$	θ	$M^2 - 2M + 1$

If $(a + 2\beta + 2\gamma + \delta + 2\eta + \theta) = 1$

$$\frac{\Sigma\delta_x\delta_y}{n} = M^2 - M(\beta + 2\gamma + \delta + 3\eta + 2\theta) + \tfrac{1}{4}\delta + \eta + \theta$$

$$= \tfrac{1}{4}\delta + \eta + \theta - M^2$$

$$\sigma_x{}^2 = \sigma_y{}^2 = \tfrac{1}{4}(\beta + \delta + 4\gamma + 5\eta + 4\theta) - M^2$$

Thus the value of r is

$$\frac{\delta + 4\eta + 4\theta - 4M^2}{\beta + \delta + 4\gamma + 5\eta + 4\theta - 4M^2}$$

For the table representing the autosomal case of the parent-child relation on page 100:

$$M = \tfrac{1}{2}(a^2b + ab) + b^3 + \tfrac{3}{2}ab^2$$
$$= b$$
$$\delta + 4\eta + 4\theta = b(1 + 3b)$$
$$\beta + \delta + 4\gamma + 5\eta + 4\theta = 2b(1 + b)$$

Hence

$$r = \frac{b(1 + 3b) - 4b^2}{2b(1 + 2b) - 4b^2} = \frac{1}{2}$$

When dominance is complete, we may rewrite the table on page 101 in the following form:

	o	$\tfrac{1}{2}$	1
o	a^3	o	a^2b
$\tfrac{1}{2}$	o	o	o
1	a^2b	o	$b(1 + ab)$

Proceeding as before,

$$M = b(1 + a)$$
$$\delta + 4\eta + 4\theta = 4b(1 + ab)$$
$$\beta + \delta + 4\gamma + 5\eta + 4\theta = 4b(1 + a)$$

When, as stated (page 101), r reduces to

$$\frac{a}{1 + a}$$

I

If we assign the representative measurements 0, A and 1 to R, H and D, the value of M is

$$A(\beta + \delta + \eta) + (\gamma + \eta + \theta)$$

And the value of r is

$$\frac{\delta A^2 + 2\eta A + \theta - M^2}{A^2(\beta + \delta + \eta) + (\gamma + \eta + \theta) - M^2}$$

APPENDIX V (page 106)

(i) When the sample is a single pair, we employ the formula for small samples, viz. $\dfrac{\Sigma\delta^2}{n-1}$. In this case $n - 1 = 1$.

The mean is

$\frac{1}{2}(x + y)$ and $\Sigma\delta^2 = [x - \frac{1}{2}(x + y)]^2 + [y - \frac{1}{2}(x + y)]^2$

So the mean square deviation for a single pair is $\dfrac{(x - y)^2}{2}$. For N pairs the mean is $\dfrac{\Sigma(x - y)^2}{2N}$.

(ii) If the correlation table is made by double entry,

$$\sigma_x = \sigma_y$$

and

$$\Sigma(x - y) = 0$$

Using the relation given as a footnote on page 67:

$$\sigma_{x-y}{}^2 = \frac{\Sigma(x - y)^2}{N} - \left[\frac{\Sigma(x - y)}{n}\right]^2$$

$$= \frac{\Sigma(x - y)^2}{N}$$

By the relation for the standard deviation of a difference given in elementary text-books of statistics,

$$\sigma_{x-y}{}^2 = \sigma_x{}^2 + \sigma_y{}^2 - 2r\sigma_x\sigma_y$$

$$= 2\sigma_x{}^2(1 - r)$$

APPENDIX VI (page 78)

In his valuable book, *Inherited Diseases of the Skin*, Dr. Cockayne tabulates all the data concerning ichthyosis congenita. Of all sibships, 24 per cent are offspring of first-cousin unions. There are in all 140 affected individuals. The expectation is $141 \cdot 1 \pm 6 \cdot 7$. For xeroderma, Siemens' summary is given. This, unfortunately, omits the size of the sibships with single affected individuals; but Cockayne has collected details of 31 sibships with family data published since the appearance of Siemens' paper. First-cousin unions are known to have occurred among 26 per cent of parents of these sibships, in which 55 affected individuals are recorded. The expectation is $53 \cdot 1 \pm 5 \cdot 0$. In both these cases the correspondence between hypothesis and observation is as close as we have any reason to expect.

APPENDIX VII

Since the problem of determining linkage relations in human beings is not a very simple one, space does not permit to recall the elementary genetic principles underlying linkage in animals and plants, as set forth in several readily accessible expositions such as those by Morgan, Punnett, and Crew. In matings which involve two linked genes, crossing-over only affects the proportion of the several phenotypes, when one of the parents is heterozygous for both. The other parent may be homozygous for both recessive genes, homozygous for one recessive gene and heterozygous for the other, or heterozygous with respect to both. The relation between the crossing-over value and phenotypic proportions in the three cases is as follows.

When one parent is heterozygous with respect to both linked genes and the other parent is the double recessive type the matings may be:

$$\text{(i)} \quad \frac{A\,|\,a}{B\,|\,b} \quad \times \quad \frac{a\,|\,a}{b\,|\,b} \qquad \text{or (ii)} \quad \frac{A\,|\,a}{b\,|\,B} \quad \times \quad \frac{a\,|\,a}{b\,|\,b}$$

If the crossing-over value expressed as a fraction of unity is denoted by c, the phenotypic proportions in the two cases are:

	AB	ab	A	B	
(i)	$\frac{1}{2}(1-c)$	$\frac{1}{2}(1-c)$	$\frac{1}{2}c$	$\frac{1}{2}c$	. . . (1)
(ii)	$\frac{1}{2}c$	$\frac{1}{2}c$	$\frac{1}{2}(1-c)$	$\frac{1}{2}(1-c)$	. . (2)

When the other parent is heterozygous with respect to one but not the other gene substitution, the phenotypic proportions can be deduced by a chess-board schema. Two types of matings occur:

$$\text{(i)} \quad \frac{A\,|\,a}{B\,|\,b} \quad \times \quad \frac{A\,|\,a}{b\,|\,b} \qquad \text{or (ii)} \quad \frac{A\,|\,a}{b\,|\,B} \quad \times \quad \frac{A\,|\,a}{b\,|\,b}$$

The first may be represented this:

Gametes	$\frac{1}{2}(1-c)AB$	$\frac{1}{2}(1-c)ab$	$\frac{1}{2}cAb$	$\frac{1}{2}caB$
$\frac{1}{2}Ab$	$\frac{1}{4}(1-c)AB$	$\frac{1}{4}(1-c)A$	$\frac{1}{4}cA$	$\frac{1}{4}cAB$
$\frac{1}{2}ab$	$\frac{1}{4}(1-c)AB$	$\frac{1}{4}(1-c)ab$	$\frac{1}{4}cA$	$\frac{1}{4}cB$

This gives the phenotypic proportions:

$$\begin{array}{ccccc}
& AB & ab & A & B \\
\text{(i)} & \frac{1}{4}(2-c) & \frac{1}{4}(1-c) & \frac{1}{4}(1+c) & \frac{1}{4}c
\end{array} \qquad . \quad . \quad . \quad (3)$$

Similarly

$$\begin{array}{cccc}
\text{(ii)} & \frac{1}{4}(1+c) & \frac{1}{4}c & \frac{1}{4}(2-c) & \frac{1}{4}(1-c)
\end{array} \qquad . \quad . \quad (4)$$

When both parents are heterozygous with respect to both gene substitutions, three matings are possible:

$$\text{(i)} \quad \dfrac{Aa}{Bb} \times \dfrac{Aa}{Bb} \qquad \text{(ii)} \quad \dfrac{Aa}{Bb} \times \dfrac{Aa}{bB} \qquad \text{(iii)} \quad \dfrac{Aa}{bB} \times \dfrac{Aa}{bB}$$

Proceeding in the same way as before, the phenotypic proportions are found to be:

$$\begin{array}{ccccc}
& AB & ab & A & B \\
\text{(i)} & \frac{1}{4}(3-2c+c^2) & \frac{1}{4}(1-c)^2 & \frac{1}{4}(2c-c^2) & \frac{1}{4}(2c-c^2) & . \;(5) \\
\text{(ii)} & \frac{1}{4}(2+c-c^2) & \frac{1}{4}c(1-c) & \frac{1}{4}(1-c+c^2) & \frac{1}{4}(1-c+c^2) & (6) \\
\text{(iii)} & \frac{1}{4}(2+c^2) & \frac{1}{4}c^2 & \frac{1}{4}(1-c^2) & \frac{1}{4}(1-c^2 & . \;(7)
\end{array}$$

In experiments with animals or plants the geneticist can arrange his procedure so as to distinguish between parents of the type

$$\text{(i)} \quad \dfrac{A \,|\, a}{B \,|\, b} \qquad \text{and (ii)} \quad \dfrac{A \,|\, a}{b \,|\, B}$$

Although the two types are not visibly dissimilar, we know that an offspring of a cross between the double recessive and the pure wild type belongs to the first, and offspring of a cross between one mutant type and the other belong to the second. In general our knowledge of the ancestry of a human being who is heterozygous with respect to both of two linked genes is not sufficient to tell us whether both recessive genes are on the same chromosome or not. In addition we only know that an individual parent is heterozygous with respect

to both genes, when it has offspring of both recessive types or of the double recessive type. There are thus two difficulties to surmount. The first arises because we cannot distinguish between two types of double heterozygotes. The second, which involves issues like those discussed on pages 73–75, arises because our census of matings involving a parent heterozygous with respect to two linked genes is necessarily incomplete. It is necessarily incomplete because of the possibility of confusing a parent of the class $Aabb$ with $AAbb$ and a parent $AaBb$ with a parent $AABB$, $AaBB$, or $AABb$. Bernstein's treatment will be easier to follow if we consider each difficulty in turn.

The second issue only arises because of the small size of the human family. If human families were very large there would be no danger of confusing matings such as those specified in the preceding remarks. We shall therefore neglect the small size of the human family to begin with, assuming that we know that a given parent is of the constitution $AaBb$, but do not know whether a and b are on the same or different members of a pair of chromosomes. Evidently we have to look for some numerical characteristic common to matings involving either

$$\text{(i)} \quad \begin{matrix} A & a \\ B & b \end{matrix} \qquad \text{or (ii)} \quad \begin{matrix} A & a \\ b & B \end{matrix}$$

If it is to tell us anything about linkage, such a characteristic must also depend upon the value of c (crossing-over value). In a cross between the double recessive and dihybrids of either type the expected proportions of individuals belonging to the two classes AB or ab on the one hand and A or B on the other are interchangeable; being $1 - c$ and c respectively when the dihybrid parent is of type (i) and c and $1 - c$ respectively when the dihybrid parent belongs to type (ii). Thus the product of the probability that and individual will belong to one of the classes AB or ab and the probability that an individual will belong to one of the classes A or B is the same in both cases; and these probabilities are themselves functions of c. Bernstein uses the sum of the observed products to evaluate c.

The problem of determining the mean value of the product of individuals belonging to one or other of the classes $(AB + ab)$ and $(A + B)$ is algebraically analogous to determining the mean value

of the product of the number of heads and the number of tails in a single set of n tosses with a coin. If the probability of getting heads in a single toss is p and the probability of getting tails is $q = (1 - p)$, the probability of getting r heads and $n - r$ tails in a set of n tosses is

$$^nC_r p^r q^{n-r}$$

The mean number of heads obtained in a single set of n tosses is

$$\Sigma r \cdot {}^nC_r p^r q^{n-r} = \underline{np} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The mean value for the product of the number of heads and the number of tails is likewise:

$$\Sigma r \cdot (n - r)^n C_r p^r q^{n-}$$
$$= \Sigma n \cdot r \cdot {}^{n-1}C_r p^r q^{n-r}$$
$$= nq\Sigma r^{n-1}C_r p^r q^{(n-1)-r}$$
$$= nq(n - 1) \cdot p \quad \text{or} \quad \underline{n(n - 1)pq} \quad . \quad . \quad . \quad (9)$$

So if we select at random from all matings of the type $AaBb \times aabb$ a fraternity of s members, the expected value $(\mu\nu)^0$ of the product formed by multiplying the number of members belonging to the classes $(AB + ab)$ and the members belonging to the classes $(A + B)$ is

$$s(s - 1)pq$$

In this expression p is the probability that an offspring of such parentage will be of the type AB or ab, and q is the probability that an offspring of such parentage will be of the type A or B. When the heterozygous parent is of type (i) $p = 1 - c$ and $q = c$, and when it is of type (ii) $p = c$ and $q = 1 - c$. Thus for matings $AaBb \times aabb$:

$$(\mu\nu)^0 = s(s - 1)c(1 - c) \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Similarly for matings of the type $AaBb \times Aabb$ or $AaBb \times aaBb$

$$(\mu\nu)^0 = s(s - 1) \cdot \tfrac{1}{4}(1 + 2c) \cdot \tfrac{1}{4}(3 - 2c) \quad . \quad . \quad (11)$$

For matings of the class $AaBb \times AaBb$ the phenotypic proportions of the three possible kinds are defined by (5), (6) and (7). In this case we may assume, as is true if equilibrium has been established after continued mating at random, that types (i) and (ii) of the

double heterozygote occur with equal frequency, and that therefore matings defined by (6) occur twice as often as matings defined by either (5) or (7). The weighted mean product is therefore:

$$(\mu\nu)^0 = s(s-1) \cdot \frac{1}{4}\left[\frac{(1+c^2)(7-c^2)}{4}\right.$$
$$\left. +\frac{(1+c-c^2)(1-c+c^2)}{2} + \frac{(2-2c+c^2)(2c-c^2)}{4}\right] \quad (12)$$

Our first difficulty is now disposed of. If all the families at our disposal constituted a perfectly random sample, we should construct tables of $(\mu\nu)^0$ for different values of s and of c as defined by (10), (11) and (12) appropriate to matings

$$AaBb \times aabb$$
$$AaBb \times Aabb \text{ or } aaBb$$
$$AaBb \times AaBb$$

On classifying the phenotypes in each family as μ members belonging to AB or ab and ν members belonging to A or B, and adding the products $\Sigma(\mu\nu)$, we could compare the sum with that of the sum of the expected products $\Sigma(\mu\nu)^0$ obtained from the tables. The variance of this quantity is readily obtained by a simple transformation from the binomial series in terms of s and c. If the quantity $\Sigma(\mu\nu)$ differed from $\Sigma(\mu\nu)^0$ by an amount more than three times its standard error for values of $(\mu\nu)^0$ calculated on the assumption that $c = \frac{1}{2}$, i.e. for an independent assortment, it would be permissible to deduce that linkage existed and to make an estimate of the linkage value by comparing $\Sigma(\mu\nu)$ with $\Sigma(\mu\nu)^0$ calculated separately for different values of c.

The reader will perhaps find it more easy to follow the next step in the analysis if we first illustrate the meaning of formula (9) with the aid of an arithmetical example. Suppose we consider 2-membered fraternities with parents $AaBb \times Aabb$. Ten different kinds of such families may be recognised according to the frequency with which each of the four phenotypes AB, ab, A and B represented therein. The composition of these ten types is given in columns (2) to (5) of the ensuing table. According to whether the parent $AaBb$ has (i) both dominant genes on one chromosome or (ii) one dominant gene on one chromosome and the other on its fellow, the probabilities

that an individual offspring will belong to each of these four classes may be denoted (i) d, e, f, g, and (ii) h, i, j, k. When $c = \frac{1}{4}$ from (3), $d = \frac{7}{16}$, $e = \frac{3}{16}$, $f = \frac{5}{16}$, $g = \frac{1}{16}$, and from (4) $h = \frac{5}{16}$, $i = \frac{1}{16}$, $j = \frac{7}{16}$, $k = \frac{3}{16}$. For the two classes of parent the frequencies of the several kinds of families are exhibited in algebraic form in column (1) and in arithmetical form in columns (7) and (9), the former being obtained by expanding $(d + e + f + g)^2$ and $(h + i + j + k)^2$. For each type of family μ is the sum of the terms in columns (2) and (3), and ν is the sum of the numbers in (4) and (5).

Frequency (n_s)	μ		ν		$\mu\nu$	(i)		(ii)	
	AB	ab	Ab	aB		n_s	$n_s\mu\nu$	n_s	$n_s\mu\nu$
d^2 or h^2	2	0	0	0	0	49	0	25	0
$2de$ or $2hi$	1	1	0	0	0	42	0	10	0
$2df$ or $2hj$	1	0	1	0	1	70	70	70	70
$2dg$ or $2hj$	1	0	0	1	1	14	14	30	30
e^2 or i^2	0	2	0	0	0	9	0	1	0
$2ef$ or $2ij$	0	1	1	0	1	30	30	14	14
$2eg$ or $2ik$	0	1	0	1	1	6	6	6	6
f^2 or j^2	0	0	2	0	0	25	0	49	0
$2fg$ or $2jk$	0	0	1	1	0	10	0	42	0
g^2 or k^2	0	0	0	2	0	1	0	9	0
Total ..	—	—	—	—	—	256	120	256	120

For the entire group of families the mean value of $\mu\nu$ is $120 \div 256$ or $0\cdot469$ in agreement with the formula given above. Thus

$$s(s -)(d + e)(f + g) = 2 \cdot 1 \cdot \tfrac{5}{8} \cdot \tfrac{3}{8} = 0\cdot469 \quad . \quad . \quad (13)$$

In testing human data we cannot generally distinguish a heterozygote from a pure dominant. Hence our sample is not a representative one. It does not include any families of which the parents are indeterminate. As the human family is small, families which are indeterminate form a considerable proportion of those produced by matings of the several types which have been considered above. Our second difficulty is to take into account the small size of the human family. In doing so we shall only consider what is involved

when the product $(AB + ab)(A + B)$ is the same whether the parent $AaBb$ is

$$\text{(i)} \quad \begin{array}{c|c} A & a \\ B & b \end{array} \qquad \text{or (ii)} \quad \begin{array}{c|c} A & a \\ b & B \end{array}$$

In other words, we shall only consider matings $AaBb \times aabb$ and $AaBb \times Aabb$.

We cannot tell whether a parent $AaBb$ is heterozygous for A, if all the offspring are AB or A, or that it is heterozygous for B, if all the offspring are AB or B. Hence we have to exclude the following types of family in practice:

(*a*) families of AB alone;
(*b*) families of A alone;
(*c*) families which consist of at least one A and at least one AB, but contain no B nor ab;
(*d*) families of B alone;
(*e*) families which consist of at least one B and at least one AB, but contain no A nor ab.

Let us revert to the arithmetical example already given in the preceding table before undertaking a more general treatment of the problem. The tabulated frequencies of the families which we reject in practice because one or both parents are indeterminate is as follows:

Frequency (n_s)	μ		ν		$\mu\nu$	(i)		(ii)	
	AB	ab	A	B		n_s	$n_s\mu\nu$	n_s	$n_s\mu\nu$
d^2 or h^2	2	0	0	0	0	49	0	25	0
$2df$ or $2hj$	1	0	1	0	1	70	70	70	70
$2dg$ or $2hk$	1	0	0	1	1	14	14	30	30
f^2 or j^2	0	0	2	0	0	25	0	49	0
g^2 or k^2	0	0	0	2	0	1	0	9	0
Total ..	—	—	—	—	—	159	84	183	100

Thus out of 256 families of class (i) and 256 families of class (ii), or 512 in all, 159 + 183 or 342 in all are excluded. Likewise, from the total value of the product $\mu\nu$, i.e. 120 + 120 or 240 in all, we

have to take away $84 + 100$ or 184. The new mean value of $\mu\nu$ is therefore

$$(240 - 184) \div (512 - 342) = 0.329$$

The mean value of the product $\mu\nu$ is

$$\frac{\Sigma(\mu\nu)n_s}{\Sigma n_s} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

In this expression n_s is the frequency of each possible class of family included in the sample examined. When the whole sample is available, as we assumed in deriving formula (9), Σn_s is $(p + q)^s$, which is unity by definition. When the sample does not include all possible families the denominator is no longer unity. Let us derive the denominator first of all. When A and B are both present on the same chromosome of the parent $AaBb$, the probability that a family will belong to one of the excluded classes (a), (b) or (c) specified above is the probability that all its members will be AB or A, i.e. it is $(d + f)^s$. Likewise the probability that any family will belong to the classes (d), and (e) is the probability that all its members will be either AB or B, excluding the families of which all the members are AB. This is $(d + g)^s - d^s$. Thus, if a parent $AaBb$ belongs to type (i) the probability of excluding it from our census of families is

$$(d + f)^s + (d + g)^s - d^s$$

If the probability that a parent $AaBb$ belongs to type (i) is $\frac{1}{2}$, the probability of excluding a family involving a parent $AaBb$ on this account alone is

$$\tfrac{1}{2}\{(d + f)^s + (d + g)^s - d^s\}$$

Likewise if a parent of the class $AaBb$ has the gene A on one chromosome and the gene B on its fellow, the probability of excluding a family because its parentage is indeterminate is

$$\tfrac{1}{2}\{(h + j)^s + (h + k)^s - h^s\}$$

From (1), (2) and (3), (4) it will be seen that the proportions of the phenotype classes is inverted in crosses involving one or the other type of parent having the constitution $AaBb$. Hence $h = f$, $i = g$, $j = d$, $k = e$. So if, as in a system which has reached equilibrium after many generations of random mating is necessarily true, both

kinds of double heterozygotes are equally common, the probability of excluding a family with one parent of the class *AaBb* and the other of the class *aabb* or *Aabb* is

$$\tfrac{1}{2}\{2(d+f)^s + (d+g)^s + (e+f)^s - d^s - f^s\}$$

Thus the probability that a family will not be excluded, that is to say the total frequency of all observed families expressed as a fraction of unity, is

$$1 - \tfrac{1}{2}\{2(d+e)^s + (d+f)^s + (e+f^s\} + \tfrac{1}{2}(d^s+f^s) = \Sigma n_s \quad . \quad (15)$$

For a family of *s* members the actual numbers belonging to the four phenotype classes may be denoted

$$
\begin{array}{cccc}
AB & ab & A & B \\
\alpha & \beta & \gamma & \delta \\
\underbrace{\hspace{2cm}}_{\mu} & & \underbrace{\hspace{2cm}}_{\nu} &
\end{array}
$$

All possible families of *s* members are derived by assigning all possible values to α, β, γ and δ from o to *s* inclusive, consistent with the relation

$$\alpha + \beta + \gamma + \delta = s$$

The frequencies of each possible class of families is obtained by expanding $(d+e+f+g)^s$ or $(h+i+j+k)^s$, the general term being

$$\frac{\lfloor s}{\lfloor \alpha \rfloor \beta \lfloor \gamma \rfloor \delta} d^\alpha e^\beta f^\gamma g^\delta \quad \text{or} \quad \frac{\lfloor s}{\lfloor \alpha \rfloor \beta \lfloor \gamma \rfloor \delta} h^\alpha i^\beta j^\gamma k^\delta$$

According as the parent *AaBb* has both genes *A* and *B* on the same chromosome or *A* on one chromosome and *B* on its fellow, the total value of the products in the numerator of (14), when we include all families, may be written:

$$(i) \quad \Sigma(\alpha+\beta)(\gamma+\delta)\frac{\lfloor s}{\lfloor \alpha \rfloor \beta \lfloor \gamma \rfloor \delta} d^\alpha e^\beta f^\gamma g^\delta$$

$$\text{or} \quad (ii) \quad \Sigma(\alpha+\beta)(\gamma+\delta)\frac{\lfloor s}{\lfloor \alpha \rfloor \beta \lfloor \gamma \rfloor \delta} h^\alpha i^\beta j^\gamma k^\delta$$

If the two types of the double heterozygote are equally common, the numerator in (14) becomes, when all families are included,

$$\Sigma \tfrac{1}{2}\mu\nu\frac{\lfloor s}{\lfloor \alpha \mid \beta \mid \gamma \mid \delta}\{d^\alpha e^\beta f^\gamma g^\delta + h^\alpha i^\beta j^\gamma k^\delta\}$$

$$\text{or}\quad \Sigma \tfrac{1}{2}\mu\nu\frac{\lfloor s}{\lfloor \alpha \mid \beta \mid \gamma \mid \delta}\{d^\alpha e^\beta f^\gamma g^\delta + d^\gamma e^\delta f^\alpha g^\beta\}$$

$$= s(s-1)(d+e)(f+g) \quad \ldots \ldots \quad (16)$$

From this quantity we must subtract the excluded families which are indeterminate. For families of AB alone, μ is zero and ν is zero. For families consisting only of AB or A, the value of α is μ and of γ is ν, so that the sum of the products is

$$\Sigma \tfrac{1}{2}\mu\nu\frac{\lfloor s}{\lfloor \mu \mid \nu}d^\mu f^\nu + \tfrac{1}{2}\mu\nu\frac{\lfloor s}{\lfloor \mu \mid \nu}f^\mu d^\nu \quad \ldots \quad (17)$$

Likewise for families of AB and B alone it is

$$\Sigma \tfrac{1}{2}\mu\nu\frac{\lfloor s}{\lfloor \mu \mid \nu}d^\mu g^\nu + \tfrac{1}{2}\mu\nu\frac{\lfloor s}{\lfloor \mu \mid \nu}f^\mu e^\nu \quad \ldots \quad (18)$$

In deriving (9), we have seen that when $(p+q)=1$ we can sum a series of the form

$$\Sigma r \cdot s \cdot {}^{r+s}C_r p^r q^s$$

In (17) and (18) $(d+f)$, $(d+g)$ and $(e+f)$ are not unity, but since $s = \mu + \nu$ these series can be put in a suitable form for summation by using the identities

$$d^\mu f^\nu = (d+f)^s\left(\frac{d}{d+f}\right)^\mu\left(\frac{f}{d+f}\right)^\nu, \text{ etc.}$$

Hence it follows that

$$\Sigma\mu\nu\frac{\lfloor s}{\lfloor \mu \mid \nu}d^\mu f^\nu = (d+f)^s\Sigma\mu\nu\frac{\lfloor s}{\lfloor \mu \mid \nu}\left(\frac{d}{d+f}\right)^\mu\left(\frac{f}{d+f}\right)^\nu$$

$$= (d+f)^s \cdot s(s-1)\left(\frac{d}{d+f}\right)\left(\frac{f}{d+f}\right)$$

$$= (d+f)^{s-2} \cdot s(s-1)df$$

The sum of the products for all the excluded families is thus:

$$\tfrac{1}{2}s(s-1)\{2(d+f)^{s-2}df + (d+g)^{s-2}dg + (e+f)^{s-2}ef\} \quad (19)$$

Subtracting this expression from (16) we obtain the value of $\Sigma\mu\nu \cdot n_s$ in (14) when the indeterminate families are excluded, and dividing by the denominator Σn_s as defined by (15) we find that the mean value of the product $\mu\nu$ for a family of s members when all indeterminate families are excluded is

$$(\mu\nu)_0 = \frac{\tfrac{1}{2}s(s-1)\{2(d+e)(f+g)-2(d+f)^{s-2}df-(d+g)^{s-2}dg-(e+f)^{s-1}ef\}}{1-\tfrac{1}{2}\{2(d+f)^s+(d+g)^s+(e+f)^s\}+\tfrac{1}{2}(d^s+f^s)} \quad (20)$$

The reader will find no difficulty in satisfying himself that this formula gives the same result as that derived directly for the case when $s = 2$ and $c = \tfrac{1}{4}$. Tables may be constructed for the value the above expression assumes for different values of s and c. Since d, e, f, g are different functions of c according as we are dealing with matings $AaBb \times aabb$ or $Aabb \times Aabb$ it will be necessary to have separate tables for the two classes of matings. With the aid of such tables we can decide whether linkage between two genes exists, and assign a cross-over value by the procedure indicated on page 137.

The last expression (20) is only true in general, when the two types of parent $AaBb$ are equally frequent. In the particular case when $c = \tfrac{1}{2}$ it is valid apart from this assumption; hence the formula given may be used to test whether linkage exists in all circumstances. It would not be useful for assigning an actual value for c, if inbreeding, as shown by Haldane and Waddington, had produced a notable discrepancy between the proportions of types (i) and (ii).

INDEX

DATE OF ISSUE

This book must be returned within 3, 7, 14 days of its issue. A fine of ONE ANNA per day will be charged if the book is overdue.